



DePaul Law Review

Volume 66
Issue 1 *Fall 2016: Twenty-Sixth Annual DePaul
Law Review Symposium*

Article 5

De Facto Mandatory: A Quantitative Assessment of Reasonableness Review After Booker

Carrie Leonetti

Follow this and additional works at: <https://via.library.depaul.edu/law-review>



Part of the [Law Commons](#)

Recommended Citation

Carrie Leonetti, *De Facto Mandatory: A Quantitative Assessment of Reasonableness Review After Booker*, 66 DePaul L. Rev. (2017)

Available at: <https://via.library.depaul.edu/law-review/vol66/iss1/5>

This Article is brought to you for free and open access by the College of Law at Via Sapientiae. It has been accepted for inclusion in DePaul Law Review by an authorized editor of Via Sapientiae. For more information, please contact digitalservices@depaul.edu.

DE FACTO MANDATORY: A QUANTITATIVE ASSESSMENT OF REASONABLENESS REVIEW AFTER *BOOKER*

Carrie Leonetti*

ABSTRACT

The Supreme Court's failure to explain how the two conflicting objectives of *United States v. Booker* and *United States v. Fanfan*—for district courts to consider the unique characteristics and circumstances of individual offenders and offenses, and for appellate courts to review sentences uniformly for reasonableness—could and should be balanced has resulted in disuniformity within and among the circuits as to whether the Sixth Amendment requires them to approve virtually any sentence within the statutory range. *Gall v. United States* has emboldened some district courts, while others remain highly guideline bound, at least in part because of the “safe harbor” that a guideline sentence can give on appeal.

All of this doctrinal uncertainty has resulted in a division among federal circuit courts of appeal as to the proper allocation of discretion between trial and appellate courts. Some circuits vest an inordinate amount of discretion at the district court level, which is unreviewable in practice. At the other end of the spectrum, other circuits vest much more discretion at the appellate level, in one of two different ways: either they retain discretion to require district courts to vary from guideline sentences for defendants whose crimes they believe the guidelines to be flawed; or they retain the discretion to require district courts to impose within-guideline sentences in all but the most extraordinary cases by setting a high threshold for sentencing variances to be affirmed on appeal. The result is a patchwork of guideline sentencing in which defendants' sentences are dictated more by geography than by the Supreme Court's jurisprudence.

* Carrie Leonetti, Associate Professor and Dean's Distinguished Faculty Fellow, University of Oregon School of Law. She wishes to thank Elisabeth Waner for her research assistance on a very complicated project. She also wishes to thank the staff of Pizza Schmizza in Raleigh Hills, who never once questioned the proportionality of exchanging eight hours of free wifi for the purchase price of one Diet Coke (with unlimited refills) because she has been forcibly ejected from restaurants in the past for less.

While other commentators have offered their own angle as to this disuniformity at a doctrinal and/or normative level, this Article seeks instead to approach it from a descriptive and empirical perspective, by attempting to quantify it numerically. It presents the results of a survey of federal appellate court cases decided after *Booker* challenging the substantive reasonableness of sentences imposed by the district court. The data suggest that the enormous discretion currently given to judges has resulted in substantial geographic disparity in application of the guidelines, district to district and circuit to circuit.

I. INTRODUCTION

There is an inherent tension between the sentencing uniformity that guideline systems were designed to promote and the judges' ability to craft fair sentences for individual offenders based on their particularized circumstances. Giving too much discretion to the sentencing judge can lead to offenders receiving dissimilar sentences for similar conduct, based solely upon the ideology of the judge that they drew. Giving too little discretion to the sentencing judge can lead to offenders with dissimilar crimes and circumstances receiving the same sentence when different sentences may be more just because the mandatory sentencing scheme failed to consider certain individualized circumstances. The benefits and disadvantages to both discretionary and mandatory sentencing schemes have caused the pendulum to swing repeatedly between the two.

II. HARD CASES, BAD LAW

In 1984, Congress passed the Sentencing Reform Act (SRA).¹ The most important provisions govern the imposition of noncapital sentences in federal court² and create the United States Sentencing Commission (USSC).³ The USSC's mission was to create sentencing guidelines that reduce "unwarranted sentencing disparities" and implement the primary purposes of criminal punishment in a just and rational manner,⁴ which led to the adoption of the United States Sentencing Guidelines (USSG) in 1987.⁵

1. Sentencing Reform Act of 1984, Pub. L. No. 98-473, tit. 2, ch. 2, 98 Stat. 1987 (codified as amended in scattered sections of 18 U.S.C. and 28 U.S.C.). The SRA was enacted as part of the Comprehensive Crime Control Act of 1984, Pub. L. No. 98-473, tit. 2, 98 Stat. 1976 (codified as amended in scattered sections of 18 U.S.C. and 28 U.S.C.).

2. 18 U.S.C. § 3553 (2012).

3. 28 U.S.C. §§ 991–995 (2012).

4. *Id.*

5. See generally Kate Stith & Steve Y. Koh, *The Politics of Sentencing Reform: The Legislative History of the Federal Sentencing Guidelines*, 28 WAKE FOREST L. REV. 223, 228–29 (1993). In

Before the USSG were put into place, a wide disparity existed among the sentences given by different judges and even among sentences given by a single judge.⁶ There was virtually no substantive appellate review of sentences in federal criminal cases at that time, as long as the sentence imposed was within the applicable statutory range.⁷ “Although the appellate courts lacked the power to review . . . the length[] of sentences imposed by district courts, they retained some ability to review the *process* through which sentences were determined. The outer limits of the district court’s discretion were set by concepts of due process.”⁸ The historical underpinnings of the USSC and the guidelines appeared more than a decade before the enactment of the SRA when, in 1973, Marvin E. Frankel published *Criminal Sentences: Law Without Order*.⁹ Judge Frankel proposed three key

the 1970s and 1980s, conservatives concerned about what they perceived as too-light sentences from lenient judges and liberals concerned about sentencing disparities joined forces in a sentencing-reform effort that ultimately became guideline sentencing. See JUSTICE KENNEDY COMM’N, ABA, REPORT WITH RECOMMENDATIONS TO THE ABA HOUSE OF DELEGATES 14–16 (2004) [hereinafter KENNEDY COMM’N REPORT], http://www.americanbar.org/content/dam/aba/publishing/criminal_justice_section_newsletter/crimjust_kennedy_JusticeKennedyCommissionReports.authcheckdam.pdf.

6. See David Fisher, *Fifth Amendment—Prosecutorial Discretion Not Absolute: Constitutional Limits on Decision Not to File Substantial Assistance Motions*, 83 J. CRIM. L. & CRIMINOLOGY 744, 745 (1993) (“Prior to the passage of the Sentencing Reform Act, federal judges enjoyed extremely broad discretion in sentencing. A judge could impose any sentence she thought was proper as long as it did not exceed the statutory maximum.”).

7. *Id.* (noting that before the SRA, there was no appellate review of sentencing decisions); Stith & Koh, *supra* note 5, at 226 (“For over two hundred years, there was virtually no appellate review of the trial judge’s exercise of sentencing discretion.”); see e.g., *Solem v. Helm*, 463 U.S. 277, 290 & n.16 (1983) (“[I]t is not the role of an appellate court to substitute its judgment for that of the sentencing court as to the appropriateness of a particular sentence.”); *Dorszynski v. United States*, 418 U.S. 424, 431 (1974) (reiterating the “general proposition that once it is determined that a sentence is within the limitations set forth in the statute under which it is imposed, appellate review is at an end”); *Herron v. United States*, 551 F.2d 62, 64 (5th Cir. 1977) (“The severity of a sentence imposed within the statutory limits will not be reviewed.”); *United States v. Cavazos*, 530 F.2d 4, 5 (5th Cir. 1976) (refusing to engage in a substantive review of Cavazos’s sentence because it was within the applicable statutory range). See generally *Appellate Review of Sentences: A Symposium at the Judicial Conference of the United States Court of Appeals for the Second Circuit*, 32 F.R.D. 249 (1962).

8. KENNEDY COMM’N REPORT, *supra* note 5, at 14 n.17; see, e.g., *United States v. Tucker*, 404 U.S. 443, 447–49 (1972) (vacating sentences from prior uncounseled convictions on due-process grounds); *Townsend v. Burke*, 334 U.S. 736, 741 (1948) (holding that a violation of due process occurred when a sentence was based on erroneous factual information); *United States v. Clements*, 634 F.2d 183, 186 (5th Cir. 1981) (refusing to “review the severity of a sentence imposed within statutory limits,” but “carefully scrutiniz[ing] the judicial process by which the punishment was imposed”); *United States v. Espinoza*, 481 F.2d 553, 558 (5th Cir. 1973) (“[The] discretion [of sentencing judges] is not, and has never been absolute, and while the appellate courts have little if any power to review substantively the length of sentences, it is our duty to insure that rudimentary notions of fairness are observed in the process at which the sentence is determined.” (citation omitted)).

9. See generally MARVIN E. FRANKEL, *CRIMINAL SENTENCES: LAW WITHOUT ORDER* (1973).

reforms to the federal sentencing system: (1) the creation of a permanent sentencing commission of sentencing experts; (2) the creation of a numerical guideline system; and (3) the requirement of meaningful appellate review of sentencing decisions to ensure a reasonable degree of “consistency and uniformity.”¹⁰

The USSG, which went into effect in 1987, provided detailed guidance for federal judges in the exercise of their sentencing authority.¹¹ Superimposed on the existing, typically broad, indeterminate statutory ranges of punishment was a class of binding, narrower guideline provisions that in many cases were driven by extremely detailed sentencing factors.¹² The guidelines provided that various aggravating and mitigating factors were to be given specific weights (upward or downward adjustments).¹³ The guidelines also significantly restricted judges’ ability to consider many offenders’ personal characteristics, like age, family circumstances, and substance-abuse history, instead focusing on an offender’s prior criminal record as the most important factor.¹⁴ In order to “depart” from the prescribed guideline sentence, a district judge had to either justify a departure by reference to factors specified in the Guidelines as appropriate grounds for doing so¹⁵ or certify that “there exist[ed] an aggravating or mitigating circumstance of a kind, or to a degree, not adequately taken into consideration by the Sentencing Commission in formulating the guidelines that should result in a sentence different from that described.”¹⁶

10. *Id.* at 115–19. Judge Frankel’s anti-overcriminalization and anti-discrimination proposal, however, was ultimately hijacked by law-and-order politics. See Stith & Koh, *supra* note 5, at 226 (“The introduction of parole in the federal system in 1910 signified a reduction in formal judicial authority over the duration of prison terms . . .”).

11. See U.S. SENTENCING GUIDELINES MANUAL app. § 5.2 (U.S. SENTENCING COMM’N 1987).

12. See *id.* § 2D1.1 (providing differing offense levels based on a long list of drug types and quantities).

13. See *id.* §§ 4A1.1, 4A1.3 (requiring that a defendant’s sentence be given an upward or downward adjustment for criminal history that is under- or over-represented by the criminal-history score).

14. *Id.* §§ 5H1.1–1.2; see KENNEDY COMM’N REPORT, *supra* note 5, at 35–37. Chapter 5, Part H of the guidelines lists factors that the USSC determined were “not ordinarily relevant to the determination of whether a sentence should be outside the applicable guideline range”: age; educational and vocational skills; mental and emotional conditions; physical condition; history of substance abuse; employment record; family or community ties; socio-economic status; military record; history of charitable good works; and lack of guidance as a youth. See U.S. SENTENCING GUIDELINES MANUAL §§ 5H1.1–1.12. See generally 28 U.S.C. §§ 994(d)–(e) (1988) (dictating that certain offender characteristics, like socioeconomic status and family ties, were generally inappropriate considerations).

15. See U.S. SENTENCING GUIDELINES MANUAL §§ 5H1.1–1.12 (enumerating permissible grounds for departure from a guideline sentence).

16. 18 U.S.C. § 3553(b)(1) (1988); U.S. SENTENCING GUIDELINES MANUAL § 5K2.0.

In 2005, the pendulum swung back. In the consolidated cases of *United States v. Booker* and *United States v. Fanfan*,¹⁷ the Supreme Court's holding on the constitutionality of the guidelines proclaimed that, in some circumstances, judicial fact-finding in connection with mandatory sentencing guidelines violated the Sixth Amendment's jury-trial guarantee.¹⁸ The solution that the Court selected in *Booker* to remedy the Sixth Amendment problem, however, was to split the baby—attempting to retain the uniform guidance of the sentencing guidelines while increasing the discretion that trial courts had to impose non-guideline sentences in appropriate cases.¹⁹ The resulting remedial holding of *Booker* rendered the USSG advisory only.²⁰ While the old mandatory guidelines had required judges to sentence within a narrow guideline range under all but the most extraordinary enumerated circumstances, the new advisory guidelines permit judges to consider the unique characteristics and circumstances of both the

17. 543 U.S. 220, 245, 268 (2005) (holding that the mandatory USSG system violated the Sixth Amendment and instructing that district courts should read the guidelines as “effectively advisory” and that “reasonableness” was the standard controlling appellate review of the sentences that district courts imposed).

18. *Id.* at 232. There were two opinions for the Court in *Booker*: (1) Justice Stevens' opinion on behalf of five justices holding that a mandatory guideline system under which a defendant's sentence would be increased on the basis of facts not proven to a jury beyond a reasonable doubt violated the Sixth Amendment (the “constitutional opinion”), *id.* at 234, and (2) Justice Breyer's opinion on behalf of five justices (only Justice Ginsburg joining both blocs), which held that the remedy for the defect identified in the constitutional opinion was to sever the portion of the SRA that made the guidelines mandatory, thus rendering them “advisory” (the “remedial” opinion), *id.* at 244–46.

19. *Id.* at 245–46.

20. *Id.* at 245. Prior to *Booker*, except when an offender cooperated with the Government against other suspects, departures pursuant to USSG § 5K2.0 represented the primary vehicle by which one party or another could seek relief from a guideline sentence. *Id.* at 234. USSG § 5K2.0 permitted district courts to depart above or below a guideline sentence when the guidelines failed to adequately take into account an aggravating or mitigating circumstance in an individual case. When it rendered the guidelines advisory, the *Booker* Court introduced the concept of a variance, which was a second, more discretionary way for district courts to impose sentences outside of the guidelines. *Id.* at 245–46. The touchstone for a variance is 18 U.S.C. § 3553(a), which authorizes courts to vary from a guideline sentence when such sentence is inconsistent with the sentencing purposes that it sets forth. As Justice Breyer observed in the remedial portion of the *Booker* opinion: “Section 3553(a) . . . sets forth numerous factors that guide sentencing. Those factors in turn will guide appellate courts . . . in determining whether a sentence is unreasonable.” *Id.* at 261.

In light of their authority to grant both departures and variances, sentencing judges are now required to engage in a multi-step, sequential sentencing process, which includes a calculation of the applicable guideline range, the consideration of any guideline-based departures, and then finally a consideration of any variance pursuant to § 3553(a). While the guidelines do not necessarily reflect all of the § 3553(a) factors, as a practical matter, most variances have a departure analogue in the guidelines. U.S. SENTENCING GUIDELINES MANUAL §§ 5K1.1–3.1 (U.S. SENTENCING COMM'N 2016).

individual offender and offense committed.²¹ With this discretion judges may adjust the severity of sentences accordingly. At the same time, however, to achieve the sentencing uniformity intended by Congress,²² the Court authorized circuit courts to review sentences for reasonableness.²³ The Court's failure to explain how these two conflicting objectives could and should be balanced (or even to acknowledge the conflict) has resulted in disuniformity within and among the circuits. The lack of uniformity lies in, as Justice Scalia predicted in his dissent from the Court's remedial opinion in *Booker*, whether the Sixth Amendment requires courts to approve "virtually any sentence within the statutory range that the sentencing court imposes, so long as the district judge goes through the appropriate formalities, such as expressing his consideration of and disagreement with the Guidelines sentence."²⁴

After *Booker*, the federal courts of appeal quickly split on the appropriate standard of review for sentences on appeal. In particular, courts differed on whether an in-guideline sentence enjoyed, on review, a presumption of reasonableness.²⁵ In 2007, the Court compounded rather than resolved this doctrinal confusion when it decided *Rita v. United States*,²⁶ which permitted but did not require appellate

21. Compare U.S. SENTENCING GUIDELINES MANUAL § 5.2 (U.S. SENTENCING COMM'N 1987), with U.S. SENTENCING GUIDELINES MANUAL § 5K2.0 (U.S. SENTENCING COMM'N 2016).

22. "The movement towards more determinate sentencing was pursued for the purpose of decreasing or eliminating the judicial discretion that many believed was the primary cause of vast sentencing disparities that were frequently based on race or class." KENNEDY COMM'N REPORT, *supra* note 5, at 59. "It was believed that the reduction or elimination of judicial discretion would result in similarly situated individuals receiving the same sentence." *Id.* In enacting the SRA, Congress sought to achieve this noble purpose of reducing "unwarranted sentencing disparities among defendants with similar records who [had] been found guilty of similar criminal conduct while maintaining sufficient flexibility to permit individualized sentences when warranted by mitigating or aggravating factors." 28 U.S.C. § 991(b)(1)(B) (2012).

23. See *Booker*, 543 U.S. at 260–61.

24. *Id.* at 312 (Scalia, J., dissenting in part) (expressing the view that the Court's remedial opinion, in an attempt to rescue from nullification a statutory scheme designed to eliminate discretionary sentencing, discarded the provisions that eliminated discretionary sentencing and did not even pretend to honor the principle that sentencing discretion was unreviewable except pursuant to specific statutory direction).

25. Compare *United States v. Dorcely*, 454 F.3d 366, 376 (D.C. Cir. 2006) (adopting a presumption of reasonableness), *United States v. Kristl*, 437 F.3d 1050, 1053–54 (10th Cir. 2006) (per curiam), *United States v. Williams*, 436 F.3d 706, 708 (6th Cir. 2006), *United States v. Green*, 436 F.3d 449, 457 (4th Cir. 2006), *United States v. Alonzo*, 435 F.3d 551, 554 (5th Cir. 2006), *United States v. Mykytiuk*, 415 F.3d 606, 608 (7th Cir. 2005), and *United States v. Lincoln*, 413 F.3d 716, 717 (8th Cir. 2005), with *United States v. Carty*, 520 F.3d 984, 994 (9th Cir. 2008) (declining to adopt a presumption of reasonableness), *United States v. Fernandez*, 443 F.3d 19, 27 (2d Cir. 2006), *United States v. Jimenez-Beltre*, 440 F.3d 514, 518 (1st Cir. 2006) (en banc), *United States v. Cooper*, 437 F.3d 324, 331 (3rd Cir. 2006), and *United States v. Talley*, 431 F.3d 784, 788 (11th Cir. 2005) (per curiam).

26. 551 U.S. 338 (2007).

courts to apply a presumption of reasonableness when reviewing in-guideline sentences.²⁷ According to the *Rita* Court, this presumption of reasonableness was “not binding,” as it was an “appellate court presumption” that applied “only on appellate review.”²⁸ The Court concluded that a presumption of reasonableness was permissible because in-guideline sentences reflected a coincidence of two independent and reasoned judgments: (1) the United States Sentencing Commission having considered the sentence in the abstract and (2) the district judge having considered the sentence in a particular case.²⁹ The *Rita* Court did not address the standard of review that an appellate court should apply when reviewing a sentence that varied from the applicable guideline range.

The following term, the plot thickened when the Court decided the companion cases of *Gall v. United States*³⁰ and *Kimbrough v. United States*.³¹ Together, these cases established that the Sixth Amendment permitted but did not require district courts to vary from a particular guideline range at sentencing when they disagreed with the policy underlying it.³² In *Kimbrough*, the Court left open the possibility of a substantive challenge to an in-guideline sentence and reminded district courts of the limited role that the guidelines were supposed to play at sentencing after *Booker*,³³ instructing that the guidelines, “for-

27. *Id.* at 362–63 (holding that the United States Court of Appeals for the Fourth Circuit properly applied a presumption of reasonableness to *Rita*’s in-guideline sentence because the presumption was not binding, did not reflect strong judicial deference, and merely recognized that a sentencing court’s discretionary sentence that was also in accord with the advisory guidelines was probably reasonable).

28. *Id.* at 347, 351. The sentencing court, however, does not enjoy the benefit of this presumption. *Id.* at 351. The Court made clear that this presumption of reasonableness was an appellate one, which should not bear on a district court’s decision. *Id.*

29. *See id.* at 350–51 (“[T]he courts of appeals’ ‘reasonableness’ presumption, rather than having independent legal effect, simply recognizes the real-world circumstance that when the judge’s discretionary decision accords with the Commission’s view of the appropriate application of § 3553(a) in the mine run of cases, it is probable that the sentence is reasonable.”).

30. 552 U.S. 38 (2007).

31. 552 U.S. 85 (2007).

32. *Gall*, 552 U.S. at 47 (rejecting the Government’s argument that district courts could only sentence outside of a guideline range in an “extraordinary” case and holding that the federal courts of appeal were required to review all federal criminal sentences under a deferential abuse-of-discretion standard); *Kimbrough*, 552 U.S. at 91, 96 (holding that the 100:1 sentencing equivalency for crack and powder cocaine was advisory, like all of the other guidelines under *Booker*, and the sentencing court did not abuse its discretion in considering the resulting sentencing disparity in imposing a sentence below *Kimbrough*’s applicable guideline range).

33. *Id.* at 101 (“[A]s a general matter, ‘courts may vary [from Guidelines ranges] based solely on policy considerations, including disagreements with the Guidelines’” (alteration in original) (quoting Brief for United States at 16, *Kimbrough*, 552 U.S. at 85 (No. 06-6330), 2007 WL 2461473, at *16)).

merly mandatory, now serve as one factor among several courts must consider in determining an appropriate sentence.”³⁴

In *Gall*, the Court affirmed that 18 U.S.C. § 3553(a) was the ultimate sentencing touchstone and that the substantive reasonableness of a sentence should be measured against its factors, noting that the guideline range was simply the “starting point and the initial benchmark” for sentencing and that district courts had to consider all of the § 3553(a) factors to determine the appropriate sentence.³⁵ The *Gall* court established that the substantive reasonableness of a sentence was to be reviewed for abuse of discretion, regardless of whether the sentence was “inside, just outside, or significantly outside” of the applicable guideline range.³⁶

In the wake of these decisions, there has been a great deal of confusion and uncertainty about whether there remains any meaningful appellate review of in-guideline sentences.³⁷ The effect of these decisions has never been measured empirically.

The recent case of Ahmed Ressim, the “Millenium Bomber,”³⁸ is a good example of how this confusion stemming from *Booker* and subsequent Supreme Court decisions play out in practice. The judge for

34. *Id.* at 90.

35. *Gall*, 552 U.S. at 49–50.

36. *Id.* at 41. *Gall* was convicted of conspiring to distribute narcotics in an enterprise from which he had withdrawn approximately three and a half years earlier. *See id.* at 41–42. His applicable guideline range was 30–37 months imprisonment, but the district court sentenced him to 36 months probation, finding that his sentence reflected the seriousness of the offense and that imprisonment was unnecessary given both his voluntary withdrawal from the conspiracy and his post-offense conduct, which included earning a college degree and starting his own successful business. *See id.* at 42–43.

The United States Court of Appeals for the Eighth Circuit reversed *Gall*’s sentence, holding that a sentence outside of the guideline range had to be supported by a justification that was “proportional to the extent of the difference between the advisory range and the sentence imposed.” *Id.* at 45 (quoting *United States v. Clairborne*, 439 F.3d 479, 481 (8th Cir. 2006)). According to the Eighth Circuit, the sentence of probation reflected an “extraordinary” 100% downward variance, which was justified by neither *Gall*’s voluntary withdrawal from the conspiracy nor his youth at the time of the offense. *Id.* Finding that the extraordinary variance was not supported by “extraordinary circumstances,” the court reversed. *Id.*

The Supreme Court, finding that the Eighth Circuit had failed to give due deference to the district court’s sentencing decision, reversed and upheld the original sentence. *Gall*, 552 U.S. at 41. The Court held that appellate courts could neither require that “extraordinary” circumstances justify a variance nor employ rigid mathematical formulas, such as the percentage of the variance, to determine whether the sentencing court’s justifications were sufficiently compelling. *Id.* at 47. Rather, even though the size of a variance was a “relevant” consideration, courts of appeal must review the final sentence imposed under a deferential abuse of discretion standard. *Id.* at 41.

37. *See* Carissa Byrne Hessick & F. Andrew Hessick, *Appellate Review of Sentencing Decisions*, 60 ALA. L. REV. 1, 3–4 (2008) (noting “the confusion that the Court’s sentencing review cases has created” since *Booker*).

38. *United States v. Ressim*, 679 F.3d 1069 (9th Cir. 2012) (en banc).

the Central District of California sentenced Ressay to a 22 year term of imprisonment for plotting to detonate explosives at Los Angeles International Airport on December 31, 1999.³⁹ The sentencing guidelines called for a sentence of approximately 65 years in prison.⁴⁰ A three-judge panel of the United States Court of Appeals for the Ninth Circuit reversed the sentence, reasoning that Ressay's below-guideline sentence was substantively unreasonable because he had backed out of an agreement to cooperate with federal investigators.⁴¹ The panel also ordered that the district judge who had imposed the below-guideline sentence be removed on remand.⁴² An en banc panel of the Ninth Circuit, in a 7-4 vote, agreed that the sentence was substantively unreasonable because it did not properly account for the severity of Ressay's crimes, which, if successful, would have killed countless people, and the en banc panel remanded the case to the same district judge for resentencing.⁴³ The dissenting judges objected that the court had overstepped its authority by overturning the sentence instead of deferring to the lower court: "Our courts are well equipped to treat each offense and offender individually, and we should not create special sentencing rules and procedures for terrorists."⁴⁴

"Most modern theories of punishment attach considerable importance to the concept of proportionality—that more serious crimes warrant more punitive sanctions than less serious crimes, and that comparably serious crimes warrant comparably severe sanctions."⁴⁵ The USSG were promulgated in an effort to increase sentencing uniformity.⁴⁶ A common criticism of the post-*Booker* sentencing regime is that advisory guidelines with weak appellate review increases un-

39. See Ian Lovett, *Appeals Court Overturns Millennium Bomb-Plot Sentence; Calls It Too Light*, N.Y. TIMES (Mar. 13, 2012), <http://www.nytimes.com/2012/03/13/us/appeals-court-overturns-millennium-bomb-plot-sentence.html>.

40. *Id.*

41. *Id.*

42. *Id.*

43. *Id.*

44. *Id.* (quoting *United States v. Ressay*, 679 F.3d 1069, 1106 (9th Cir. 2012) (en banc) (Schroeder, J., dissenting)).

45. Michael Tonry, *Rethinking Unthinkable Punishment Policies in America*, 46 UCLA L. REV. 1751, 1752 (1999); see, e.g., H.L.A. HART, PUNISHMENT AND RESPONSIBILITY 3, 9, 23–24 (1968) (arguing that a punishment system that regularly imposed disproportionate punishments would defy conventional morality, create a sense of public insecurity, and lose credibility); ANDREW VON HIRSCH, CENSURE AND SANCTIONS 6, 15–19 (1993) (arguing that proportionality is a central feature of any system of punishment); NORVAL MORRIS, THE FUTURE OF IMPRISONMENT 11, 26–27 (1974) (arguing that proportionality is one valid concern among many and sets outer limits on morally justifiable punishments).

46. See *supra* notes 1–16 and accompanying text.

warranted sentencing disparities.⁴⁷ *Gall* has emboldened some district courts to exercise greater discretion, while others remain highly guideline bound. This is at least in part because of the “safe harbor” that a guideline sentence can give on appeal. Justice Souter, in his dissenting opinion in *Rita*, expressed his concern with the “gravitational pull” of the guidelines and the extent to which it could be strengthened by a presumption of reasonableness:

What works on appeal determines what works at trial, and if the Sentencing Commission’s views are as weighty as the Court says they are, a trial judge will find it far easier to make the appropriate findings and sentence within the appropriate Guideline, than to go through the unorthodox factfinding necessary to justify a sentence outside the Guidelines range.⁴⁸

In addition, as other commentators have noted, there is a tension between sentencing discretion at the district court level and substantive-reasonableness-review discretion at the circuit court level that is unresolved, and this Article contends that it remains wholly unresolvable under the Supreme Court’s current Sixth Amendment jurisprudence.⁴⁹

All of this doctrinal uncertainty has resulted in a division among federal circuit courts of appeal as to the proper allocation of discretion between trial and appellate courts. Some circuits vest an inordinate amount of discretion at the district court level, which is unreviewable in practice. At the other end of the spectrum, some circuits vest much more discretion at the appellate level. Circuits that retain appellate-level discretion do so in two different ways: Either they retain discretion to require district courts to vary from guideline sentences concerning those crimes for which they believe the guidelines to be flawed, or they retain discretion to set a high threshold for affirming

47. See, e.g., U.S. SENTENCING COMM’N, DEMOGRAPHIC DIFFERENCES IN FEDERAL SENTENCING PRACTICES: AN UPDATE OF THE BOOKER REPORT’S MULTIVARIATE REGRESSION ANALYSIS 15 (2010); see also Douglas A. Berman, *Exploring the Theory, Policy, and Practice of Fixing the Broken Sentencing Guidelines*, 21 FED. SENT’G REP. 182, 182 (2009) (criticizing the guidelines’ reliance on drug quantity and monetary loss as proxies for culpability); Ryan W. Scott, *The Effects of Booker on Inter-Judge Sentencing Disparity*, 22 FED. SENT’G REP. 104, 104 (2009).

48. *Rita v. United States*, 551 U.S. 338, 390–91 (2007) (Souter, J., dissenting) (citations omitted).

49. The Court’s two objectives in *Booker*, requiring district court discretion and cabining that discretion through reasonableness review, are in tension with each other. “[A]ppellate review, by its nature, tends to restrict the discretion of district courts in future cases and, therefore, threatens to undermine the district court discretion necessary to *Booker*’s remedy.” Hessick & Hessick, *supra* note 37, at 3. As a result, the Court has, in subsequent cases, sacrificed the central functions of appellate review: error correction and lawmaking. It has undermined the error-correction function by permitting appellate courts to presume that within-guideline sentences are reasonable, and it has impaired the lawmaking function by directing appellate courts to defer to district courts’ sentencing policy determinations.

sentencing variances and implicitly require district courts to impose in-guideline sentences in all but the most extraordinary cases. The result is a patchwork of guideline sentencing in which defendants' sentences are dictated more by the happenstance of geography than by the Supreme Court's jurisprudence.

The United States Courts of Appeal for the First, Third, Eighth, and Eleventh Circuits exemplify the latter. One example of the high threshold approach taken by these circuits is *United States v. D'Amico*.⁵⁰ D'Amico was a city councilor convicted of extortion and false statements in violation of 18 U.S.C. §§ 1951 and 1001, respectively.⁵¹ His applicable guideline range was 31–44 months imprisonment, but the district court varied downward and sentenced him to 4 months imprisonment.⁵² The variance was granted based on, among other things, the collateral consequences of D'Amico's conviction, which would likely preclude his ability to engage in professional work, and his prior good works as a legislator.⁵³ On appeal, the First Circuit vacated D'Amico's sentence, finding that the district court's reasons for the variance were insufficient.⁵⁴ The First Circuit noted that the variance resulted in a sentence 88% below D'Amico's guideline range.⁵⁵ Although the court did not disagree with the district court's finding that D'Amico had performed prior good works as a public servant, it concluded that the sentencing court had "overvalued these contributions in imposing [D'Amico's] sentence" and had erred in considering the collateral consequences as a justification for a substantial variance.⁵⁶

A second example is *United States v. Tomko*.⁵⁷ Tomko was the owner and president of a construction company.⁵⁸ In the late 1990s, Tomko built a luxurious new home for himself and instructed the subcontractors to bill the expenses to his company, rather than to him personally.⁵⁹ The company then took illegal tax deductions on the costs of the new home as business expenses, permitting Tomko to evade taxes of more than \$200,000, for which Tomko pled guilty to tax evasion.⁶⁰ The parties agreed, as part of the plea agreement, that

50. 496 F.3d 95 (1st Cir. 2007).

51. *Id.* at 98.

52. *Id.*

53. *See id.* at 105–06.

54. *Id.*

55. *Id.* at 106.

56. *D'Amico*, 496 F.3d at 107.

57. *United States v. Tomko*, 498 F.3d 157 (3d Cir. 2007).

58. *Id.* at 159.

59. *Id.*

60. *Id.* at 159–60.

Tomko's guideline range for sentencing was 12–18 months imprisonment, but the district court varied downward and sentenced Tomko to a sentence of community service, a fine, and probation with a period of home detention.⁶¹ The Third Circuit vacated Tomko's sentence, disagreeing with the district court's analysis of the sentencing factors that justified a variance and the weight of the evidence.⁶² The Third Circuit held the sentence was substantively (as opposed to procedurally) unreasonable because it was inconsistent with the goals of 18 U.S.C. § 3553(a).⁶³

A third example is *United States v. Carlson*.⁶⁴ Carlson was convicted of willfully failing to account for and pay taxes on trust funds in violation of 26 U.S.C. § 7202.⁶⁵ His applicable guideline range was 18–24 months imprisonment.⁶⁶ The district court sentenced him to 8 months home confinement, probation, and community service.⁶⁷ The Eighth Circuit affirmed the variance because:

Carlson (1) ha[d] a significant record of charitable activities; (2) accepted responsibility and made an exceptional effort to repay [the taxes that he owed]; (3) suffered damage to his business, reputation, and family relationships; and (4) was not motivated by a desire to defraud the government, but was instead attempting to resolve a financial crisis within his business.⁶⁸

On appeal, the Eighth Circuit vacated Carlson's sentence.⁶⁹ The court noted that the variance amounted to a 100% reduction below Carlson's guideline range, noting: "'We have suggested that a variance to zero prison time where the Sentencing Commission has found that substantial prison time is indicated requires extraordinary justification,' and 'have routinely rejected this kind of variance as unreasonable.'" ⁷⁰ The court held that the specific basis for the district court's downward variance did not justify its extraordinary nature.⁷¹

The decision in *Gall*, on its face, would seem to foreclose the type of review employed by circuit courts in this latter category. Nonetheless, below-guideline sentences that have been imposed in these circuits

61. *Id.* at 161–62.

62. *Id.* at 172.

63. *Tomko*, 498 F.3d at 172.

64. *United States v. Carlson*, 498 F.3d 761 (8th Cir. 2007).

65. *Id.* at 762.

66. *Id.*

67. *Id.*

68. *Id.* at 763.

69. *Id.* at 762.

70. *Carlson*, 498 F.3d at 764 (citations omitted) (quoting *United States v. Soperla*, 494 F.3d 752, 755 (8th Cir. 2007)).

71. *Id.* at 765–66.

since *Rita* have been reversed on appeal as unreasonable. In addition, deference to district courts that impose sentences with which the circuit courts disagree, has not been as strong as the *Booker* and *Rita* decisions suggested. The review in these cases often more closely resembles de novo than abuse of discretion.

A final example is *United States v. Pugh*, which was decided after *Gall*.⁷² The Eleventh Circuit reviewed the imposition of a sentence of probation after Pugh's conviction for knowingly possessing child pornography in violation of 18 U.S.C. § 2252A(a)(5)(B).⁷³ Pugh's applicable guideline range was 97–120 months, and the district court varied substantially downward to impose a sentence of probation.⁷⁴ The district court held sentencing hearings over two days, heard from numerous witnesses, and “deliberated extensively over its sentencing decision.”⁷⁵ The Eleventh Circuit nonetheless reversed, finding Pugh's sentence to be substantively unreasonable.⁷⁶ While the appellate court described its review in deferential terms, its substantive disagreement with the district court's sentencing analysis was about how to weigh sentencing factors in a particular case.⁷⁷

This Article should not be read as criticizing judicial discretion globally or advocating a return to the mandatory guideline system that the Court struck down in *Booker*. Historically, the USSG have worked to the detriment of criminal defendants generally, tending to increase their sentences and resulting in arbitrary and unequal justice.⁷⁸ While guidelines may help sentencing courts impose fair and equitable sentences, judicial discretion is necessary to assure that sentences reflect the totality of circumstances regarding an offender and offense. A key requirement for achieving justice is giving judges sufficient discretion to recognize legitimate differences among offenders charged with similar offenses and consider the totality of circumstances in deciding that a sentence in a specific case is “just.” No set of rules, including the USSG, can substitute for this necessary flexibility. After all, a system would be perfectly “uniform” if every offender were sentenced to 10 years imprisonment upon conviction for any offense, but no one would describe such system as just or fair because a fair system also requires equality.

72. *United States v. Pugh*, 515 F.3d 1179 (11th Cir. 2008).

73. *Id.* at 1182.

74. *Id.*

75. *Id.* at 1192.

76. *Id.* at 1194.

77. *See id.*

78. *See* KENNEDY COMM'N REPORT, *supra* note 5, at 16–17.

There has never been much, if any, evidence to support the view that sentencing disparity was eliminated by the mandatory guidelines.⁷⁹ On the contrary, mandatory guidelines tend to mask arbitrary disparities under the guise of methodological calculations. Mandatory guidelines fail to account for the individual when determining the appropriate punishment. From the outset, many academics criticized the guidelines as unfairly requiring consideration of “relevant uncharged conduct,” limiting consideration of relevant offender characteristics, and replacing judges’ sentencing discretion with a rigid mathematical formula.⁸⁰

The result was inappropriate uniformity for vastly different defendants and circumstances due to emphasis on a single commonality, typically the charging statute, drug quantity, or loss amount. “What proponents of determinate sentencing did not fully realize was that the elimination of judicial discretion at the sentencing stage would not eliminate disparities as long as police and prosecutors continued to exercise discretion at the arrest, charging, and plea bargaining stages of the process.”⁸¹

79. See Frank O. Bowman III, *The Quality of Mercy Must Be Restrained, and Other Lessons in Learning to Love the Federal Sentencing Guidelines*, 1996 WIS. L. REV. 679, 716 (1996) (“The somewhat troubling truth is that the actual evidence on the question of the Guidelines’ effect on disparity is scant.”); A. Abigail Payne, *Does Inter-Judge Disparity Really Matter? An Analysis of the Effects of Sentencing Reforms in Three Federal District Courts*, 17 INT’L REV. L. & ECON. 337, 338 (1997) (finding that the effect of the uniform Guidelines on the variability in sentencing was negligible); cf. David Robinson, Jr., *The Decline and Potential Collapse of Federal Guideline Sentencing*, 74 WASH. U. L.Q. 881, 912 (1996) (“Guidelines are premised on a mistaken notion that the largest problem in federal sentencing is disparity in the sentences judicially imposed on different defendants, not in erecting a system that is rational and just.”). But see James M. Anderson et al., *Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines*, 42 J.L. & ECON. 271, 294–95 (1999) (finding a decrease in inter-judge sentencing disparity after the enactment of the federal guidelines); Paul J. Hofer et al., *The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity*, 90 J. CRIM. L. & CRIMINOLOGY 239, 241 (1994) (finding “the sentencing guidelines have had modest but meaningful success at reducing unwarranted disparity among judges in the sentences imposed on similar crimes and offenders”).

80. See, e.g., William W. Schwarzer, *Judicial Discretion in Sentencing*, 3 FED. SENT’G REP. 339, 341 (1991); see also Daniel J. Freed, *Federal Sentencing in the Wake of Guidelines: Unacceptable Limits on the Discretion of Sentencers*, 101 YALE L.J. 1681, 1684–85 (1992); Marc L. Miller, *Domination and Dissatisfaction: Prosecutors as Sentencers*, 56 STAN. L. REV. 1211, 1236–37 (2004) (noting the various criticisms levied at the sentencing guidelines after their promulgation); Marc L. Miller & Ronald F. Wright, *Your Cheatin’ Heart(land): The Long Search for Administrative Sentencing Justice*, 2 BUFF. CRIM. L. REV. 723, 723–24 (1999).

81. KENNEDY COMM’N REPORT, *supra* note 5, at 109.

“[T]he elimination of judicial discretion [through determinate sentencing guidelines] strengthened the impact of the decisions made by these officials, especially the prosecutor. Judicial discretion had operated as a check on the unbridled, discretionary decisions of prosecutors, who were not otherwise accountable for their decisions. With the removal of judicial discretion and the introduction of the sentencing guidelines and

A new study has “found vast disparities in the prison sentences handed down by [district] judges presiding over similar cases.”⁸² This study by the Transactional Records Access Clearinghouse (TRAC) at Syracuse University analyzed the sentences imposed by 885 federal district judges in more than 370,000 criminal cases between the years 2007–2011 has discovered extensive variations in the sentencing practices of district judges in different districts.⁸³ Because the report examined differences in sentencing practices within individual districts where judges presumably received the same general mix of cases, this finding raises questions about the extent to which sentences in some districts are influenced by the particular judge rather than the facts of the specific case.⁸⁴

Geography is a fact neither pled nor proven to a jury beyond a reasonable doubt. When a difference in geography results in a comparatively longer sentence for some defendants than for other similarly (or even less-favorably) situated defendants in other districts, this lack of uniformity violates the Fifth and Sixth Amendments to the United States Constitution. Equality and the rule of law dictate that an offender’s sentence should not depend upon which judge happens to be assigned to the case. While myriad other commentators have offered a doctrinal and/or normative angle as to this disuniformity, this Article seeks instead to approach it from a descriptive and empirical perspective by quantifying the disuniformity.⁸⁵

Part III describes the methodology used in this survey of federal appellate court cases decided after *Booker* challenging the substantive

mandatory minimums [Congress and the USSC] essentially empowered prosecutors to predetermine the sentence through their charging and plea-bargaining decisions.”

Id.

82. Mosi Secret, *Wide Sentencing Disparity Found Among U.S. Judges*, N.Y. TIMES (Mar. 5, 2012), <http://www.nytimes.com/2012/03/06/nyregion/wide-sentencing-disparity-found-among-us-judges.html>.

83. *Id.* The study covered each sentence imposed by federal district court judges over the past five years. *Id.* See *Surprising Judge-to-Judge Variations Documented in Federal Sentencing*, TRAC REP., <http://trac.syr.edu/tracereports/judge/274> (last visited Aug. 19, 2016); see, e.g., U.S. SENTENCING COMM’N, FINAL QUARTERLY DATA REPORT tbl. 2 (2010), http://www.ussc.gov/sites/default/files/pdf/research-and-publications/federal-sentencing-statistics/quarterly-sentencing-updates/USSC_2010_Quarter_Report_Final.pdf (noting that, within the First Circuit, in-guideline sentences in the District of Massachusetts represented only 32.1% of the total cases, while in the District of Puerto Rico, in-guideline sentences represented 73% of cases).

84. See Secret, *supra* note 82. Ryan Scott reached a similar conclusion in his study of the Massachusetts federal courts. See generally Scott, *supra* note 47 (finding that sentencing disparities have increased and the effect of the individual judge has doubled since *Booker*).

85. There is no comparable data set for sentencing decisions that neither party has appealed from the district court to a court of appeals. Such decisions are typically not reported, and many defendants waive their right to appeal a court’s sentencing decision as part of a plea bargain.

reasonableness of sentences imposed by the district court.⁸⁶ Part IV presents the results of the survey.⁸⁷ The data suggest that the enormous discretion currently given to judges has resulted in substantial geographic disparity in (1) application of the guidelines, district to district and circuit to circuit, and (2) the distribution of discretion between district courts and appellate courts among the circuits. Part V offers an interpretation of these results and some concluding thoughts.⁸⁸

III. METHODOLOGY: THE METHOD TO THE MADNESS

A. *Project Design*

This project collected, collated, and coded all of the post-*Gall/Kimbrough* substantive-reasonableness appeals, published and unpublished decisions issued⁸⁹ between 2008–2011, sorted by outcome and circuit. Only cases in which the appellate court ruled on the merits of a party's substantive-reasonableness challenge were included in the data set; cases in which a party challenged the substantive reasonableness of a sentence but the sentence was reversed on other grounds (e.g., procedural errors in the sentencing),⁹⁰ without the court reaching the question of the substantive reasonableness of the sentence imposed, were not considered. The results of reasonableness review were separated into four categories: (1) in-guideline sentences that were affirmed; (2) in-guideline sentences that were reversed; (3) sentencing variances that were affirmed; and (4) sentencing variances that were reversed. The study then compared the results of substantive reasonableness review across the twelve circuits.

86. See *supra* notes 89–93 and accompanying text.

87. See *supra* notes 95–175 and accompanying text.

88. See *supra* notes 176–185 and accompanying text.

89. This was accomplished in three steps: (1) by Shephardizing *Booker* by circuit; (2) by narrowing the results by date to include only those appeals decided after the Supreme Court issued its opinions in *Gall* and *Kimbrough*; and (3) reading the resulting cache of opinions and filtering out those that did not involve a review of the substantive reasonableness of the defendant's sentence. This methodology was selected on the assumption that, while not every case citing *Booker* would involve reasonableness review, presumably almost every, if not every, case that involved reasonableness review would cite *Booker*.

90. See, e.g., *United States v. Gunter*, 462 F.3d 237, 239, 247–49 (3d Cir. 2006) (holding that the district court erred in treating crack cocaine sentencing differential incorporated in the guidelines as mandatory as opposed to discretionary); *United States v. Crosby*, 397 F.3d 103, 119–20 (2d Cir. 2005) (holding that remand is warranted when sentencing judge commits a procedural error in imposing a sentence, affording the judge an opportunity to determine whether the original sentence would have been nontrivially different post-*Booker*).

B. Terminology & Sample

In its remedial opinion in *Booker*, the Court read a “reasonableness” standard of review into the SRA.⁹¹ This project analyzed only appeals of a sentence’s *substantive*, as opposed to procedural, reasonableness.

“Variance,” for the purpose of this study, refers to any sentence outside of the applicable guideline range, either above or below. The study did not distinguish between formal departures under the guidelines and true *Booker*/§ 3553(a) variances.⁹² In other words, if the defendant’s sentence was outside of the applicable guideline range, it was a variance. This decision was made despite the fact that, strictly speaking, a presumption of reasonableness should mean that a departure under the guideline manual is presumptively reasonable, whereas a *Booker* variance is not. This decision was made for two reasons. The first was a practical one; while *Booker* requires lower courts to distinguish guideline-based departures from § 3553-based variances as part of procedural reasonableness, appellate-court opinions do not always indicate which was imposed, particularly if the appeal only involves a challenge to the substantive reasonableness of the sentence. The second was a statistical assumption that with a large enough sample, any difference in how appellate courts treat a traditional departure versus a new variance would not skew the results in the aggregate. As a practical matter, the study found no departures outside of the guideline range that were explicitly described as guideline-based departures, but rather, all appeared to have been true variances, probably in light of the more deferential standard of review that variances receive post-*Booker*.

A “case,” for the purpose of this study, is defined per defendant. For example, if three co-defendants in a drug case were sentenced within the guidelines, appealed their sentences for substantive reasonableness in a consolidated appeal, and if the circuit court had affirmed their sentences, the study would count the result as an affirmation of three in-guideline sentences.

The study did not include in its data set appeals of sentences that were driven by mandatory minimum sentences that exceeded the otherwise applicable guideline range, even when a defendant attempted to appeal those mandatory sentences on the ground that they were substantively unreasonable or violated the parsimony principle of § 3553(a). It also did not include government appeals of sentences

91. United States v. Booker, 543 U.S. 220, 261 (2005)

92. See *supra* note 9 and accompanying text.

that were imposed below a statutory mandatory minimum, whether within the applicable guideline range or not.

The study did not control for child pornography offenses, even though two federal circuits have essentially found the child pornography guidelines to be per se unreasonable.⁹³ The number of child pornography cases in the data set was relatively small, and the study makes the assumption that the federal circuits that do not apply a presumption of reasonableness to the child pornography guidelines likely do not apply a presumption of reasonableness to the guidelines generally; therefore, the circuits' caselaw regarding child pornography sentences would not be expected to skew the results.

IV. RESULTS & ANALYSIS: THE MADNESS IN THE METHOD⁹⁴

As Justice Scalia predicted in his dissent from the Court's remedial opinion in *Booker*, the majority opinion has "produce[d] a discordant symphony of different standards, varying from court to court and judge to judge."⁹⁵ Striking sentencing disparities jump out at first glance. In some districts and circuits, judges regularly depart from the guideline sentencing range for whole categories of crimes—for example, in cases involving possession and receipt of child pornography and certain types of fraud;⁹⁶ in other districts and circuits, judges generally impose in-guideline sentences only.⁹⁷ The results are sometimes drastically inconsistent across cases and among similarly situated defendants, and it raises questions about the extent to which the goal of

93. See *United States v. Grober*, 624 F.3d 592, 598 (3d Cir. 2010); *United States v. Dorvee*, 616 F.3d 174, 186–87 (2d Cir. 2010).

94. All data is on file with the author.

95. *Booker*, 543 U.S. at 312 (Scalia, J., dissenting in part).

96. This is, perhaps, not surprising given that the guidelines for both types of offenses have been widely criticized for significant recent increases to guideline sentences with no basis in empirical data, but rather in response to a series of ill-considered congressional directives passed at the urging of law enforcement agencies. See, e.g., *Osborne v. Ohio*, 495 U.S. 103, 108–10 (1990) (rejecting a First Amendment challenge to the criminalization of mere possession of child pornography on the ground that the demand for child pornography encouraged its production, which necessarily involved the physical sexual abuse of children). But see Michael C. Seto et al., *Contact Sexual Offending by Men with Online Sexual Offenses*, 23 SEXUAL ABUSE J. RES. & TREATMENT 124, 128–31, 140 (2011) (performing a meta-analysis of 33 studies that looked at the criminal histories and recidivism rates of child pornography offenders and concluding that child pornography offenders were relatively unlikely to commit sexual offenses involving physical contact with a child). In a 2010 survey by the USSC, approximately 70% of federal judges said that the recommended penalties for possessing or receiving child pornography were unreasonable. U.S. Sentencing Comm'n, RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES: JANUARY 2010 THROUGH MARCH 2010 tbls. 1 & 8 (2010), http://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-projects-and-surveys/surveys/20100608_Judge_Survey.pdf.

97. See, e.g., *United States v. Duarte*, 569 F.3d 528, 530 (5th Cir. 2009).

equal justice under the law is served by the Court's sentencing jurisprudence.

A. Reversal Data

The results of this survey are as follows. The United States Court of Appeals for the First Circuit affirmed 22 in-guideline sentences, reversed 1 in-guideline sentence, affirmed 10 variances, and reversed zero variances. The United States Court of Appeals for the Second Circuit affirmed 53 in-guideline sentences, reversed 1 in-guideline sentence, affirmed 33 variances, and reversed 7 variances. The United States Court of Appeals for the Third Circuit affirmed 71 in-guideline sentences, reversed zero in-guideline sentences, affirmed 25 variances, and reversed 11 variances. The United States Court of Appeals for the Fourth Circuit affirmed 73 in-guideline sentences,⁹⁸ reversed 5 in-guideline sentences,⁹⁹ affirmed 41 variances,¹⁰⁰ and reversed 8 variances. The United States Court of Appeals for the Fifth Circuit reviewed 172 sentences for reasonableness post-*Booker*, 93 in-guideline and 79 variances, and affirmed all but 2 of those sentences, reversing 2 in-guideline sentences in a single case (involving two co-defendants). The United States Court of Appeals for the Sixth Circuit affirmed 74 in-guideline sentences, reversed 5 in-guideline sentences, affirmed 38 variances, and reversed 7 variances. The United States Court of Appeals for the Seventh Circuit affirmed 21 in-guideline sentences,¹⁰¹ reversed 1 in-guideline sentence, affirmed 13 variances,¹⁰² and reversed 5 variances. The United States Court of Appeals for the Eighth Circuit affirmed 29 in-guideline sentences, reversed zero in-guideline sentences, affirmed 22 variances, and reversed 5 variances. The United States Court of Appeals for the Ninth Circuit affirmed 28 in-guideline sentences, reversed 6 in-guideline sentences, affirmed 16 variances, and reversed 2 variances. The United States Court of Ap-

98. This number includes 19 appeals in which the defendant's appellate counsel filed a brief pursuant to *Anders v. California*, 386 U.S. 738, 744 (1967) (requiring that a court-appointed-appellate attorney who finds a defendant's appeal to be wholly frivolous to advise the court and supply a brief referring to anything in the record that might arguably support the appeal).

99. This includes 1 appeal in which the defendant's appellate counsel filed a brief pursuant to *Anders*.

100. This includes 4 appeals in which the defendant's appellate counsel filed a brief pursuant to *Anders*.

101. This includes 2 appeals in which the defendant's appellate counsel filed a brief pursuant to *Anders*.

102. This includes 1 appeal in which the defendant's appellate counsel filed a brief pursuant to *Anders*.

peals for the Tenth Circuit affirmed 53 in-guideline sentences,¹⁰³ reversed 1 in-guideline sentence, affirmed 22 variances, and reversed 3 variances. The United States Court of Appeals for the Eleventh Circuit affirmed 157 in-guideline sentences, reversed 1 in-guideline sentence, affirmed 71 variances, and reversed 5 variances. The United States Court of Appeals for the District of Columbia Circuit affirmed 5 in-guideline sentences, reversed zero in-guideline sentences, affirmed 3 variances, and reversed 2 variances.¹⁰⁴

B. Interpretation of Reversal Data

The introduction to this Article divided the federal circuits into three categories based on their style of post-*Booker* review: (1) those that show a great deal of deference to district court sentences; (2) those that exercise a great deal of discretion at the appellate level, reversing relatively high percentages of district court sentences, but doing so with little regard to whether such sentences are within or outside the applicable guideline range and, in some instances, even requiring district courts to vary from guideline sentences when sentencing defendants for crimes for which they believe the guidelines to be flawed; and (3) those that continue to show a great deal of deference to the guidelines, requiring district courts to impose in-guideline sentences in all but the most extraordinary cases by setting a high threshold for sentencing variances to be affirmed on appeal. These categories are somewhat fluid (requiring a subjective decision as to what constitute relatively “high” and “low” rates of reversals), but the statistical disparities between both overall reversal rates and relative reversal rates as well as between in-guideline sentences and variances are so large that they warrant systematic analysis.

103. This includes 4 appeals in which the defendant’s appellate counsel filed a brief pursuant to *Anders*.

104. There were clearly significant variations in the number of appeals of sentences based on claims of substantive reasonableness, which do not correlate with the size of caseload of the respective circuit involved. For example, the Ninth Circuit is the largest circuit in the country, by geography, population, and caseload, see U.S. COURTS FOR THE NINTH CIRCUIT, 2015 NINTH CIRCUIT ANNUAL REPORT 3 (2015), but it has only reviewed 50 sentences for substantive reasonableness during the course of the study, fewer total sentences than variances reviewed by the Eleventh Circuit during the same time frame. The possible causes of this seeming disparity (e.g., rates of appeal waivers, rates of appellate reversal within the circuits, rates of below-guidelines sentences imposed in the districts within the circuits, or the level of deference that a circuit shows either the guidelines or district courts on appeal) are interesting but beyond the subject matter of this study.

1. *Deference to District Court*

The First, Fifth, Tenth, and Eleventh Circuits demonstrated extraordinary deference to district court sentences in their post-*Booker* reasonableness review with little difference between their treatment of in-guideline sentences and variances on appeal. This is best demonstrated by looking at the comparative percentages of affirmances and reversals, rather than the raw numbers. The First Circuit affirmed 97% of a total of 33 sentences on reasonableness review, affirming 22 in-guideline sentences and 10 variances, while reversing only 1 in-guideline sentence. The Fifth Circuit affirmed approximately 99% of the sentences that it reviewed. The only 2 sentences that it reversed were within the applicable guideline range.

The results for the Tenth and Eleventh Circuits were less stark, but nonetheless warrant their placement in this category, with approximately 95% and 96% affirmance rates, respectively. Each court was slightly more likely to reverse a variance than an in-guideline sentence, but not by a large enough margin to classify them as courts that showed significant deference to the guidelines. The Tenth Circuit reversed in-guideline sentences approximately 2% of the time and variances approximately 12%; the Eleventh Circuit reversed in-guideline and variance sentences at similar rates 1% and 11%, respectively, but both circuits' overall reversal rates (both less than 5%) were so low as to render the differences between the percentages for those two subtypes of reversals insignificant.

2. *Significant Discretion Within Appellate Court*

The Sixth, Seventh, and Ninth Circuits belong in an opposing category, showing significantly less deference to district court sentences on appeal and more willingness to reverse sentences that they found to be substantively unreasonable. In percentage terms, the Sixth Circuit reversed just less than 10% of the sentences that it reviewed for substantive reasonableness, reversing approximately 6% of in-guideline sentences and 16% of variances. The Seventh Circuit reversed approximately 15% of all challenged sentences, approximately 5% of in-guideline sentences and approximately 26% of variances. The Ninth Circuit reversed approximately 14% of the challenged sentences, reversing approximately 17% of in-guideline sentences and 16% of variances.¹⁰⁵

105. These findings are consistent with prior findings that the Seventh and Ninth Circuit's review of guideline departures, under the formerly mandatory guideline scheme pre-*Booker*, were anomalous and "aggressive." See Michael S. Gelacak et al., *Departures Under the Federal*

3. *Deference to Guidelines*

The Second, Third, Fourth, Eighth, and D.C. Circuits all showed significantly more deference to guideline sentences than to variances, with overall reversal rates that were relatively high, but almost entirely due to their high rates of variance reversal. While there were no circuits that showed a pre-*Booker* level of deference to the guidelines—because all circuits either affirmed at least one district court variances or reversed at least one district court in-guideline sentence, which presumably would have been disallowed or at least much less likely under the old-mandatory-guideline regime—the D.C. Circuit came the closest to matching this description. The D.C. Circuit reversed 20% of all challenged sentences, but reversed zero in-guideline sentences, reversing approximately 68% of variances.

The Second Circuit reversed approximately 9% of the 94 sentences that it reviewed, but, of the 9%, only one was an in-guideline sentence, resulting in a relative reversal rate of less than 2% for in-guideline sentences and more than 18% for variances.

The Third Circuit similarly reversed fewer than 10% of all challenged sentences, but none of its reversals were of in-guideline sentences. Its relative reversal rates of in-guideline and variance sentences were 0% and more than 25%, respectively.

The Fourth Circuit reversed approximately 9% of the sentences that it reviewed, but it reversed variances at twice the rate than in-guideline sentences, reversing in-guideline sentences approximately 6% of the time and variances approximately 12% of the time.¹⁰⁶

The Eighth Circuit reversed approximately 9% of all challenged sentences, but reversed zero in-guideline sentences while reversing approximately 19% of variances.

These circuits that show considerable deference to the district courts (largely affirming all sentences whether within the guideline range or outside) are probably the most concerning, from a constitutional point of view, because the guidelines appear to have a presump-

Sentencing Guidelines: An Empirical and Jurisprudential Analysis, 81 MINN. L. REV. 299, 337–38 (1996) (finding that “the Seventh Circuit’s departure review tended to be somewhat more aggressive than that of the other courts” and that “[t]he Ninth Circuit was also notable in its departure review standards, as the court imposed rather stringent procedural requirements on district courts’ decisions’ departure”).

106. If one considers only the reasonableness appeals in which no party filed an *Anders* brief on the assumption that an *Anders* brief is an indication that an appeal is frivolous, a difficult assumption in at least the 4 *Anders* appeals involving sentencing variances and the 1 *Anders* appeal involving the in-guideline sentence that was reversed, the Fourth Circuit reversed approximately (10% of the challenged sentences, reversing approximately 8% of challenged in-guideline sentences and 13% of challenged variances.

tion of reasonableness so strong in them that they remain de facto mandatory.¹⁰⁷

4. Commonalities

There are several trends that emerge from this reversal data common to almost all of the federal circuits. In all 12 of the circuits, the majority of district court sentences were affirmed on appeal. In all of the circuits except for the D.C. Circuit, the appeals court affirmed a majority of variances, as well. In all of the circuits except for the Ninth Circuit,¹⁰⁸ the appellate courts reversed variances by a significantly higher percentage than they reversed in-guideline sentences. In every circuit, a majority of the appealed sentences were within the applicable guideline range. Of the 1,120 cases comprising the data set of the study, 674 of the challenged sentences, 60% were in-guideline sentences. Of those guideline sentences, the circuit courts collectively reversed approximately 0.5%.

One of the fascinating aspects of the study results is that the level of deference that the federal circuits show to in-guideline sentences in practice bears no apparent relationship to whether they have adopted a presumption of guideline reasonableness as a doctrinal matter. For example, of the 5 circuits that show the most deference to in-guideline sentences, only three (the Fourth,¹⁰⁹ Eighth,¹¹⁰ and D.C.¹¹¹ Circuits) have adopted this presumption. The Second¹¹² and Third¹¹³ Circuits have explicitly rejected the presumption of guideline reasonableness, but nonetheless together reversed only 1 in-guideline sentences out of a total of 124 guideline sentences reviewed (i.e., fewer than 1%).¹¹⁴ Conversely, the Fifth Circuit purports to apply a presumption of reasonableness for guideline sentences,¹¹⁵ but, in the entire data set of

107. See *infra* Part III.E.

108. The Ninth Circuit reversed in-guideline sentences and variances at roughly equal rates, 17% and 16% of the time, respectively.

109. See, e.g., *United States v. Green*, 436 F.3d 449, 457 (4th Cir. 2006).

110. See, e.g., *United States v. Lincoln*, 413 F.3d 716, 717 (8th Cir. 2005).

111. See, e.g., *United States v. Dorcey*, 454 F.3d 366, 376 (D.C. Cir. 2006).

112. See, e.g., *United States v. Fernandez*, 443 F.3d 19, 27 (2d Cir. 2006), *abrogated by* *Rita v. United States*, 551 U.S. 338, 346 (2007) (declining to establish a presumption that a Guidelines sentence is reasonable).

113. See, e.g., *United States v. Cooper*, 437 F.3d 324, 331–32 (3d Cir. 2006), *abrogated by* *Rita*, 551 U.S. at 346.

114. In addition to rejecting a presumption of reasonableness for in-guideline sentences, the Second and Third Circuits are the two federal circuits that have held that the child pornography guidelines are presumptively unreasonable. See *supra* note 93 and accompanying text.

115. See, e.g., *United States v. Alonzo*, 435 F.3d 551, 554 (5th Cir. 2006).

this study, reversed only a single sentence that it reviewed for reasonableness, and that sentence was within the applicable guideline range.

What emerges from these commonalities is the following, perhaps unsurprising, generality about post-*Booker* sentencing in federal courts. District courts imposing in-guideline sentences are almost certain to be affirmed on appeal, even in circuits that retain to themselves a great deal of appellate discretion and do not apply a presumption of reasonableness for in-guideline sentences.¹¹⁶ Variances are also far more likely than not (but not as likely as in-guideline sentences) to be affirmed.¹¹⁷ The result of these two cardinal principles is that the combination of deference to the district courts and deference to the guidelines creates a near-total safe harbor on appeal for district courts that sentence within the guidelines.

C. *Direction of Variance*¹¹⁸

1. *Little to No Effect*

The First and Fifth Circuits affirmed all variances that they reviewed. The First Circuit affirmed 5 upward and 5 downward variances, and the Fifth Circuit affirmed 69 upward and 10 downward variances.¹¹⁹

116. See *supra* Part IV.A (discussing results of the survey of post-*Booker* review of sentences for reasonableness, by district).

117. See *supra* Part IV.B.2.

118. The direction of variances was not always ascertainable from the appellate opinion. The results reported in this section of the study are based solely on those opinions in which the direction of the variance is indicated. The results reported in Sections III.C–F do not include downward variances that were appealed by defendants as insufficient, since, in the absence of a cross-appeal by the government, the court could not reverse the variance in favor of an in-guideline sentence. See, e.g., *Greenlaw v. United States*, 554 U.S. 237, 240 (2008) (holding that appellate courts could not increase defendants' sentences that they challenged on appeal in the absence of a government cross-appeal); *United States v. Huff*, No. 09-5021, 2010 WL 3010762, at *1 (4th Cir. July 28, 2010) (appealing the reasonableness of a below-guideline sentence of fifteen months imprisonment) (per curiam); *United States v. Sellers*, Nos. 09-4332, 09-4335, 2010 WL 2340252, at *1 (4th Cir. June 10, 2010) (appeal, inter alia, of the reasonableness of a below-guideline sentence of 144 months imprisonment); *United States v. Wilson*, No. 08-4789, 2009 WL 1262401, at *1 (4th Cir. May 8, 2009) (appeal of the reasonableness of a below-guideline sentence of 320 months imprisonment) (per curiam).

119. Two Fifth Circuit cases involving review of downward variances have been omitted from these results (analyzing the effect of the direction of a variance on the court's substantive reasonableness review) because of their anomalous circumstances. In the first case, the defendant was found guilty of 13 violations of supervised release. See *United States v. Calhoun*, No. 07-51349, 2008 WL 3821315, at *1 (5th Cir. Aug. 15, 2008). His resulting sentencing-guideline range was 4–10 months imprisonment on each count. *Id.* The trial court sentenced the defendant to 2.3 months imprisonment on each count, but ran the thirteen sentences consecutively. See *id.* While the individual sentences were technically downward variances, the defendant was appealing the consecutive nature of the imposition of the sentences, which resulted in a total sentence (30

The Sixth Circuit reversed 2 out of 21 upward variances and 3 out of 24 downward variances that it reviewed, for relative reversal rates of approximately 10% and 13%.

The Eighth Circuit reversed a significant percentage of the variances that it reviewed, but the direction of the variance did not play a significant role. The court reversed approximately 19% of the downward variances (4 out of 21) and 17% of the upward variances (1 out of 6) that it reviewed.

The Ninth Circuit only reversed 1 upward variance, out of 5 total that it reviewed, and 1 downward variance, out of 13 total that it reviewed.¹²⁰ Similarly, the Tenth Circuit reversed 2 out of 16 upward variances and 1 out of 9 downward variances, for relative reversal rates of approximately 13% and 11%, respectively.¹²¹

The D.C. Circuit only reviewed 5 total variances: 3 downward and 2 upward. To the extent that any meaningful information can be gleaned from such a small data set, it does not appear that the direction of the variance affected the likely result. The court affirmed 2 downward and 1 upward variance and reversed 1 upward and downward variances each.

2. *Preference for Upward Variances*

Several circuits were significantly more likely to reverse a downward variance than an upward variance. While the downward variances (26) reviewed by the Second Circuit outnumbered the upward variances (12) by more than two to one,¹²² all 7 of the variances that

months significantly above the high end of the guideline range for the individual counts (10 months), had the sentences been run concurrently. *Id.*

In the second case, the defendant was found guilty of multiple offenses, some which carried consecutive mandatory-minimum sentences. *See United States v. Harper*, 527 F.3d 396, 399–400 (5th Cir. 2008). The trial court imposed the mandatory sentences, plus a consecutive 10-month sentence on the remaining counts, which was significantly below the applicable 63–79 month guideline range for those counts. *Id.* at 411. Harper appealed, arguing that the total sentence (370 months) was unreasonable; however, because the court only had the discretion to sentence him to a more lenient sentence on the counts that did not carry mandatory-minimum sentences, he was, in essence, appealing a downward departure from 63–10 months as insufficiently below the applicable guideline range. *Id.*

120. While statistically this creates a significant variation in the percentage of upward (8%) and downward (17%) variances that the court reversed, one of each category did not seem to be a large enough sample from which to conclude that the direction of the variance played a role in the outcome of the court's review.

121. Similar to the results for the Ninth Circuit, the total number of variances that the Tenth Circuit reviewed and reversed is too small to extrapolate anything from these relative reversal rates.

122. It was impossible to tell the direction of the other 2 of the 40 total variances that the Second Circuit reviewed.

the Second Circuit reversed were downward variances. The result is that the Second Circuit reversed 27% of downward variances, compared with 0% upward variances.

The Third Circuit reviewed 36 variances, 25 of which were downward variances and 11 of which were upward variances. Of the 11 variances that the court reversed on appeal, 10 were downward variances. The result is that the Third Circuit reversed downward variances (40% of the time) at a rate of more than four times upward variances (9%).

The Fourth Circuit reviewed 49 total variances, 26 of which were upward variances and 19 of which were downward variances.¹²³ The court reversed downward variances at almost 3 times the rate of upward variances, reversing approximately 10% of upward variances and 28% of downward variances. These results likely significantly understate the relative disparity between the Fourth Circuit's reversal of upward and downward variances, because the majority of the affirmed downward variances in the data set were appeals in which the defendant was challenging a downward variance as insufficiently below the guidelines.¹²⁴

The Eleventh Circuit affirmed upward variances at a rate of almost 10 times that at which it affirmed downward variances, affirming all but 1 of the 41 upward variances that it reviewed, a relative reversal rate of approximately 2% and reversing 7 of the 34 downward variances that it reviewed,¹²⁵ a relative reversal rate of approximately 21%.¹²⁶

123. It was impossible to ascertain the direction of 3 of the variances that the Fourth Circuit reviewed. In addition, one appellate case reviewing an upward variance has been left out of this section of the data set for this Article because the upward variance in question was imposed pursuant to a plea agreement in which the defendant agreed to an upward variance in exchange for a charge bargain, which dismissed more serious charges that would have resulted in a guideline range higher than the upward variance on the charges to which the defendant pleaded guilty. *See United States v. Brown*, No. 09-4122, 2009 WL 2843578, at *2 (4th Cir. Sept. 3, 2009) (per curiam).

124. *See, e.g., United States v. Huff*, No. 09-5021, 2010 WL 3010762, at *3 (4th Cir. July 28, 2010) (per curiam); *United States v. Sellers*, Nos. 09-4332, 09-4335, 2010 WL 2340252, at *3 (4th Cir. June 10, 2010) (per curiam); *United States v. Peribian-Gonzales*, No. 07-4460, 2009 WL 4049150, at *1-2 (4th Cir. Nov. 24, 2009) (per curiam); *United States v. McConnell*, No. 08-4773, 2009 WL 1262404, at *2 (4th Cir. May 8, 2009); *United States v. Wilson*, No. 08-4789, 2009 WL 1262401, at *2 (4th Cir. May 8, 2009) (per curiam).

125. It was impossible to ascertain the direction of one variance that the court affirmed. *See United States v. Walker*, 439 F. App'x. 822 (11th Cir. 2011).

126. Again, this is consistent with pre-*Booker* research, which documents that the Eleventh Circuit "rejected downward departure factors almost as consistently as it affirmed upward departures." Gelacak et al., *supra* note 105, at 350.

3. Preference for Downward Variances

The Seventh Circuit conversely appeared to show a preference for downward variances. The court reviewed 18 total variances, 11 of which were upward variances and 7 of which were downward variances. The court reversed upward at almost three times the rate of downward, reversing approximately 13% of downward variances and 33% of upward variances.¹²⁷

D. Size of Variance (Percentage Terms)¹²⁸

1. Methodology

For the purpose of this analysis, the survey divided variances into percentiles based on the relationship between the variant sentence imposed and the nearest guideline sentence to the variant sentence, by comparing upward variances to the high end of the applicable guideline range and downward variances to the low end and reporting them in 10 percent increments (0–10%, 10–20%, etc., of the applicable guideline sentence).¹²⁹ For the purpose of calculating the size of variances, the study treated any sentence other than imprisonment (house arrest, community corrections center, residential treatment center, probation/supervised release) as a sentence of zero months imprisonment.¹³⁰ In the case of both upward and downward variances, the percentage of variation was calculated by dividing the sentence imposed (the variance) by the nearest guideline sentence (high or low) and then subtracting 100% (or 1.0) from the resulting percentage.¹³¹ For example, if a court imposed a sentence of 90 months on an individual

127. This trend appears to be the opposite of that found by pre-*Booker* research, which documents that the “Seventh Circuit consistently rejected downward departures” under the mandatory guideline scheme. *Id.* at 344.

128. The size of variances was not always ascertainable from the appellate opinion. The results reported in this section of the study are based solely on those opinions in which the size of the variance is indicated.

129. The results in this section do not include downward variances from guidelines calling for a sentence of life imprisonment or upward variances to life imprisonment from guidelines calling for a sentence of a term of imprisonment, as there was no methodologically sound way to quantify a term of years as a percentage of life without parole. When a court imposed a downward variance to some form of supervision (house arrest, probation, etc.) from a guideline range calling for imprisonment, the variance was treated as a sentence of zero months imprisonment (i.e., a 100% downward variance from the low end of the guideline range).

130. The study also treated a sentence of twelve-months-plus-one-day imprisonment as a 12-month sentence, even though, as a practical matter, a defendant sentenced to twelve-months-plus-one-day in federal prison will likely serve a shorter sentence than a prisoner sentenced to 11–12 months in prison because of the eligibility for early release for good conduct. *See* 18 U.S.C. § 3624(b) (2012).

131. As a mathematical result, downward variances cannot exceed 100%, but upward variances can because judges cannot impose sentences of negative years imprisonment.

whose low end of the applicable guideline range was 100 months, the resulting sentence would reflect a downward variance of 10%. If a court imposed a sentence of 110 months on an individual whose high end of the applicable guideline range was 100 months, the resulting sentence would reflect an upward variance of 10%. The percentiles were calculated so that a variance of exactly 10% counted in the 10–20% percentile, exactly 20% counted in the 20–30% percentile, etc. (i.e., the percentiles were really $10\% < 20\%$, etc.).

2. *Size Matters*

The size of a downward variance, by percentage, appeared to play a significant role in the likelihood that the Second, Eighth, and Eleventh Circuits would reverse it as substantively unreasonable. The Second Circuit reviewed 12 downward variances¹³² in amounts ranging from 0–39% and affirmed all but 1 of them, a relative reversal rate of approximately 8%. The court reviewed 8 downward variances in amounts between 40–99% and reversed 5 of them, a relative reversal rate of 63%. The court reviewed 2 downward variances of 100%, variances to residential treatment and probation from low guideline ranges of 12 and 108 months imprisonment, respectively, and reversed both of them, a relative reversal rate of 100%.

The Eighth Circuit reviewed 24 total variances during the study period. It affirmed all 8 of the variances of less than 40% and affirmed 11 of the 16 variances that it reviewed of more than 40%, for relative reversal rates of zero for small variances and 31% for larger ones.

At first glance, the Eleventh Circuit appears to favor variances that are very large or very small, only reversing “medium-sized” variances at a significant rate. The court reversed only 1 of the 19 variances that were less than 30% of the applicable guideline range and 1 of the 29 variances that were more than 60% of the applicable guideline range, but reversed 6 of 25 variances in amounts between 30% and 60% of the applicable guideline range. The resulting relative reversal rates create a visual donut effect, with small variances having a relative reversal rate of approximately 5%, large variances having a relative reversal rate of approximately 3%, and only medium-size variances having a relative reversal rate of 24%. While it is possible that the circuit truly shows some preference for medium-sized variances,¹³³ the

132. These results do not include upward variances because the Second Circuit affirmed 100% of the upward variances it reviewed, including 6 in which the imposed sentence exceeded the high end of the guideline range by more than 100%.

133. While one would expect courts generally to affirm small variances at a higher rate than larger ones, the counter-intuitive result of the Eleventh Circuit data is the relatively lower rever-

apparent donut effect appears to be accounted for by two other variables (besides the size of the variance): (1) the circuit's preference for upward variances over downward ones, considering only upward variances are mathematically able to exceed 100% of the applicable guideline range,¹³⁴ and (2) the small number of downward variances of more than 30% below the low end of the applicable guideline range that district courts in the Eleventh Circuit imposed in the first instance. When one analyzes only downward variances, the "donut" disappears, with the size of the variance playing a more predictable (and significant) role in the likelihood that a variant sentence would be affirmed on appeal. The court reversed 1 out of 15 downward variances that were less than 30% below the low end of the guideline range, 4 out of 13 downward variances that were between 30–60% below the low end of the guideline range, and 1 of the only 2 total downward variances that were between 60–100% below the low end of the guideline range. This results in relative reversal rates of less than 7% for small downward variances, approximately 31% for medium-sized downward variances, and 50% for large downward variances.

The size of the variance, by percentage, appeared to play some role, albeit not as significant, in the likelihood that the Third Circuit would reverse the sentence as substantively unreasonable. The court reviewed 18 variances that were less than 50% above or below the respective guidelines and reversed 4 of them, for a relative reversal rate of less than 30%.¹³⁵ The court reviewed 5 variances that were between 50–100% above or below the respective guidelines and reversed 2 of them, for a relative reversal rate of 40%. The court reviewed 5 variances of 100% or more above or below the respective guidelines and reversed 3 of them, for a relative reversal rate of 60%.

The size of the variance in the Seventh Circuit, by percentage, appeared to play a small role, although the total number of variances was so small in the Seventh Circuit that the results may not be statistically significant. The court reversed 2 out of 13 variances of less than 50% below or above the applicable guidelines, 2 out of 4 variances of

sal rate of very high variances in comparison to medium-sized ones. One possible explanation for a significant decline in the rate at which variances are reversed when they grow from somewhat sizeable to very large could be that district courts are more careful about making a record for the reason for variances as those variances become larger.

134. See *supra* note 126 and accompanying text.

135. These results omit one anomalous case in which the Third Circuit reversed a downward variance of 40% below the applicable guideline sentence because it was insufficient (i.e., the below-guideline sentence was unreasonably high). See *United States v. Olhovsky*, 562 F.3d 530, 532–33 (3d Cir. 2009).

50–100% of the applicable guidelines, and 1 out of 2 variances of more than 100% of the applicable high end of the guideline range, for a relative reversal rate of approximately 15% for variances of less than 50% and 50% for variances of more than 50%.

The size of the variance, by percentage, appeared to play some role in the Fourth, Sixth, and Tenth Circuits, but only at the extremes. The Fourth Circuit affirmed 19 of the 21 variances of less than 100%, but 6 of the 10 variances of more than 100%, for relative reversal rates of less than 10% and 60%, respectively.¹³⁶ This is particularly noteworthy given that these variances in excess of 100%, under the methodology employed to reach these results, could only be upward variances and that the Fourth Circuit generally showed a significantly higher likelihood of reversing downward variances than upward ones.¹³⁷

Similarly, the Sixth Circuit affirmed 32 of the 35 variances of less than 100%, but reversed 3 of the 8 variances of more than 100%, for relative reversal rates of less than 9% and 38% respectively.

The Tenth Circuit affirmed 19 of the 20 variances of less than 200% but 2 of the 4 variances of more than 200% that it reviewed, for relative reversal rates of 5% and 50%. The one less than 200% variance that it reversed was a downward variance of approximately 62%.¹³⁸

3. *No Apparent Impact*

The First Circuit and Fifth Circuits affirmed all of the variances that they reviewed, apparently without regard to the size of the variance in relation to the underlying guideline range. The greatest upward variance that the First Circuit affirmed, by percentage, was an upward variance of 227%, affirming a 75-month sentence of imprisonment when the applicable guideline range was 27–33 months.¹³⁹ The greatest downward variance that it affirmed, by percentage, was a downward variance of 95%, affirming a 3-month sentence of imprisonment when the applicable guideline range was 63–78 months imprisonment.¹⁴⁰

The greatest upward variance that the Fifth Circuit affirmed, by percentage, was an upward variance of 567%, affirming a 60-month sentence of imprisonment when the applicable guideline range was

136. These results omit one upward variance that was imposed pursuant to a plea agreement that included a charge bargain to avoid a higher mandatory-minimum sentence. *See United States v. Brown*, No. 09-4122, 2009 WL 2843578, at *1–2 (4th Cir. Sept. 3, 2009) (per curiam).

137. *See supra* Part III.C.

138. *See United States v. Friedman*, 554 F.3d 1301, 1302 (10th Cir. 2009).

139. *See United States v. Brown*, 631 F.3d 573, 580 (1st Cir. 2011).

140. *See United States v. Thurston*, 544 F.3d 22, 23–24 (1st Cir. 2008).

3–9 months.¹⁴¹ The greatest downward variance that the Fifth Circuit affirmed, by percentage, was a downward variance of 100%, affirming a sentence of probation when the applicable guideline range was 27–33 months imprisonment.¹⁴²

The Ninth Circuit affirmed all but 1 of the 151 variances, an upward variance of 33%.¹⁴³ The highest variance that the Ninth Circuit affirmed was an upward variance of 414%.¹⁴⁴

4. *Inconclusive*

The data set in the D.C. Circuit is too small from which to draw meaningful conclusions about the effect that the size of the variance, by percentage, had on the court's likelihood to reverse or affirm.¹⁴⁵

E. *Size of Variance (Number of Months)*

1. *Methodology*

For the purpose of this analysis, the survey divided variances into categories based on the total size of the variance by months; comparing upward variances to the high end of the applicable guideline range and downward variances to the low end and reporting them in decile increments (0–10 months, 10–20 months, etc., above or below the applicable guideline sentence).¹⁴⁶ For example, if a court imposed a sentence of 90 months on an individual whose low end of the applicable

141. See *United States v. Simmons*, No. 07-60646, 2008 WL 1897718, at *1 (5th Cir. Apr. 30, 2008) (per curiam).

142. See *United States v. Duhon*, 541 F.3d 391, 394, 399 (5th Cir. 2008).

143. See *United States v. Burgum*, 633 F.3d 810, 811 (9th Cir. 2011).

144. See *United States v. Fitch*, 659 F.3d 788, 790 (9th Cir. 2011).

145. The D.C. Circuit only reviewed 4 variances whose values could be computed as a percentage of the applicable guideline range. See, e.g., *United States v. Mejia*, 597 F.3d 1329, 1332, 1334 (D.C. Cir. 2010) (affirming a sentence of 208 months imprisonment when Mejia's applicable guideline range called for 10 years to life imprisonment). It reversed two of the variances: one upward variance of 100%, see *In re Sealed Case*, 527 F.3d 188, 190 (D.C. Cir. 2008) (reversing the district court's upward variance to a term of 18 months imprisonment from a guideline range of 3–9 months), and one downward variance of 20%, see *In re Sealed Case*, 552 F.3d 841, 842, 844 (D.C. Cir. 2009) (reversing the district court's downward variance to a term of 97 months imprisonment from a guideline range of 121–51 months). It affirmed the other two variances whose value could be calculated, an upward variance of 105%, see *United States v. Ventura*, 650 F.3d 746, 747–48 (D.C. Cir. 2011) (affirming Ventura's 84-month sentence when his applicable guideline range was 33–41 months), and a downward variance of 100%, see *United States v. Gardellini*, 545 F.3d 1089, 1090 (D.C. Cir. 2008) (affirming Gardellini's sentence of probation when his applicable guideline range was 10–16 months imprisonment).

146. The results in this section do not include downward variances from guidelines calling for a sentence of life imprisonment or upward variances to life imprisonment from guidelines calling for a sentence of a term of imprisonment because there was no methodologically sound way to quantify a term of years as a percentage of life without parole. When a court imposed a downward variance to some form of supervision (house arrest, probation, etc.) from a guideline range

guideline range was 100 months, the resulting sentence would reflect a downward variance of 10 months. The deciles were calculated so that a variance of 10 months counted in the 10–20 month decile, 20 months counted in the 20–30 month decile, etc. (i.e., the deciles were really 10 months to less than 20 months, etc.).

2. *Size Still Matters*

Like the size of the variance in percentage terms, the size of the variance in terms of total months also played a role in the Second, Seventh, Eighth, and Eleventh Circuits. The Second Circuit reviewed 14 total variances of less than 30 months and affirmed all but one of them, a relative reversal rate of approximately 7%. The 7 of them, a relative reversal rate of approximately 37%.

The Seventh Circuit reviewed 12 variances of fewer than 40 months and reversed 2 of them, for a relative reversal rate of 17%. The court reviewed 7 variances of 50 or more months and reversed 3 of them,¹⁴⁷ for a relative reversal rate of 43%.

The pattern for the Eighth Circuit appeared at the extremes, although that is likely true because the court reviewed only 1 variance of between 50 and 80 months.¹⁴⁸ The court reviewed 18 variances of fewer than 80 months and reversed 2 of them, for a relative reversal rate of 11%. The court reviewed 6 variances of 80 or more months and reversed 3 of them, for a relative reversal rate of 50%.¹⁴⁹

The Tenth Circuit appeared to be highly sensitive to the size of the total variance in months. Of the 25 variances that the court reviewed, it reversed only the 3 largest: variances of 94, 135, and 159 months, respectively.¹⁵⁰

calling for imprisonment, the variance was treated as a sentence of zero months imprisonment (i.e., a 100% downward variance from the low end of the guideline range).

147. The court did not review any variances between 40 and 50 months.

148. See *United States v. O'Connor*, 567 F.3d 395, 396, 398 (8th Cir. 2009) (affirming a 120-month sentence when O'Connor's advisory guideline range was 188–235 months imprisonment).

149. These data are not clear-cut. When one separates *large* variances by months (80–100 months) from *very large* variances by months (more than 200 months), the court reversed 100% of the large ones and affirmed 100% of the very large ones.

150. See *United States v. Lente*, 647 F.3d 1021, 1023–24 (10th Cir. 2011) (reversing Lente's 192-month sentence on remand); *United States v. Friedman*, 554 F.3d 1301, 1302 (10th Cir. 2009) (reversing Friedman's 57 month sentence when his applicable guideline range was 151–88 months imprisonment); *United States v. Lente*, 2009 WL 1143167, at *1 (10th Cir. Apr. 29, 2009) (per curiam) (reversing Lente's 216-month sentence when his applicable guideline range was 46–57 months imprisonment), *abrogated by* *United States v. Story*, 635 F.3d 1241 (10th Cir. 2011) (holding that increasing incarceration to promote rehabilitative goals is not permitted under § 3582(a)).

The size of the variance, by month, had a clearer correlation than the size of the variance, by percentage, with the likelihood that the Eleventh Circuit would reverse. Though the court again reversed fairly large variances (40–70%) at an even higher rate than exceptionally large variances (80% or more). This pattern seems to be significant in light of the fact that it holds over the size of variances in both percentage and raw-month terms and may be accounted for by the fact that upward variances are much more capable of being extraordinarily large in size than downward variances (since a court cannot impose a sentence of negative time). The Eleventh Circuit reviewed 33 variances of less than 30 months and reversed only 1 of them, resulting in a relative reversal rate for small variances of approximately 3%. The court reviewed 42 variances of 30 months or more and reversed 7 of them, for a relative reversal rate of approximately 17%. The court reversed neither of the 2 sentences of more than 200 months (the two extraordinarily large variances).

3. *No Apparent Impact*

The First and Fifth Circuits affirmed all of the variances that they reviewed, apparently without regard to the size of the variance in relation to the underlying guideline range.¹⁵¹ The greatest upward variance that the First Circuit affirmed, by months, was an upward variance of 49 months, affirming a 240-month sentence of imprisonment for a defendant whose advisory guideline range was 177–191 months.¹⁵² The greatest downward variance that it affirmed, by months, was a downward variance of 91 months, affirming a 144-month sentence of imprisonment for a defendant whose advisory guideline range was 235–93 months.¹⁵³ The greatest upward variance that the Fifth Circuit affirmed, by months, was a whopping 460-month upward variance, affirming a 1,082-month sentence imposed upon a defendant whose advisory guideline range was (an apparently insufficient) 360–622 months imprisonment.¹⁵⁴ The greatest downward variance that is affirmed was a 120-month downward variance, affirming a 240 month sentence for a defendant whose advisory guideline range was 360 months to life imprisonment.¹⁵⁵

151. See *supra* note 119 and accompanying text.

152. See *United States v. Vega-Santiago*, 519 F.3d 1, 2, 6 (1st Cir. 2008) (en banc).

153. See *United States v. Martin*, 520 F.3d 87, 88, 98 (1st Cir. 2008).

154. See *United States v. Walters*, No. 08-51223, 2009 WL 2903705, at *1 (5th Cir. Sept. 11, 2009) (per curiam).

155. See *United States v. Murray*, 648 F.3d 251, 252–53 (5th Cir. 2011).

Unlike when analyzed by percentage, when analyzed by number of months, the size of the variance did not seem to affect the likelihood that it would be affirmed or reversed in the Third, Fourth, or Sixth Circuits.¹⁵⁶ The Third Circuit reviewed 23 variances of fewer than 50 months and affirmed 17 of them, for a relative reversal rate of approximately 26%. The court reviewed 10 variances of 50 months or more and affirmed 7 of them, including 4 sentences with variances of more than 100 months, for a relative reversal rate of approximately 30%.

The Fourth Circuit reviewed 26 variances of fewer than 50 months and reversed 6 of them, for a relative reversal rate of approximately 23%. The Fourth Circuit reviewed 14 variances of more than 50 months and reversed 4 of them, for a relative reversal rate of 29%.¹⁵⁷

The Sixth Circuit reviewed 28 variances of fewer than 50 months and reversed 5 of them, for a relative reversal rate of 19%. The court reviewed 15 variances of 50 or more months and reversed only 1 of them, for a relative reversal rate of 7%. If anything, the total size of the variance by months inversely affected the likelihood of its reversal in the Sixth Circuit.¹⁵⁸

4. *Inconclusive*

The Ninth Circuit reviewed 8 variances of fewer than 50 months and reversed 1 of them, for a relative reversal rate of 13%. The court reviewed 5 variances of more than 50 months and reversed 1 of them, for a relative reversal rate of 20%.¹⁵⁹

156. This finding seems intuitive. The “size” of a variance seems more meaningful by percentage than by raw months (i.e., a 24-month variance downward from a 30-month guideline seems like a meaningful commentary of the disproportionality of the guidelines when applied to the individual defendant’s case while a twenty-four-month variance downward from a 360-month guideline seems like a token concession).

157. These data are somewhat skewed by a single, multiple co-defendant case, in which the Fourth Circuit reversed variances of 119, 203, and 222 months, respectively. See *United States v. Monroe*, Nos. 08-5050, 08-5051, 08-5052, 2010 WL 3721524, at *6 (4th Cir. Sept. 20, 2010) (per curiam).

158. This finding is likely statistically anomalous in light of the positive effect that variances had, by percentage, on reversal in the circuit and the relatively few large variances, by month, that the court reviewed.

159. While, standing alone, 20% is significantly larger than 13%, it would be hard to draw a conclusion that the size of the variance, by months, affected the Ninth Circuit’s likelihood of reversal in light of the small number (2) of total variances that it reversed, 1 of 45% and 1 of 516%. At best, one might be able to extrapolate that the Ninth Circuit favors very small variances (fewer than 40 months), but the size of the data set probably does not allow even for that conclusion.

The data set in the D.C. Circuit is too small from which to draw meaningful conclusions about the effect that the size of the variance in months, had on the court's likelihood to reverse or affirm it.¹⁶⁰

F. Case Type

1. Methodology

This Article categorized the cases in which variances were reviewed for reasonableness by broad case types and analyzed the five most common types of cases that occurred in each circuit¹⁶¹ to see if there were any correlations between the type of crime involved and the likelihood that a circuit court would affirm or reverse a variance (and in what direction). The types of cases analyzed fell into the following categories: drugs, guns, child pornography, immigration, white-collar (e.g., bank and wire fraud), bank robbery, and violent person crimes (e.g., murder).¹⁶² The purpose of this analysis was to ascertain if any circuits were demonstrating a systematic approval or disapproval of the guideline ranges for particular types of offenses in their relative treatment of variances from the applicable guidelines on appeal.

2. No Apparent Impact

The First Circuit and Fifth Circuits affirmed all of the variances that they reviewed, apparently without regard to the underlying crime charged.¹⁶³

160. Of the 4 variances whose values could be computed by months that the D.C. Circuit reviewed, it reversed 2 of the variances, 1 upward variance of 9 months, *see In re Sealed Case*, 527 F.3d 188, 192–93 (D.C. Cir. 2008), and 1 downward variance of 24 months, *see In re Sealed Case*, 552 F.3d 841, 842–44 (D.C. Cir. 2009). It affirmed the other 2 variances whose value could be calculated, an upward variance of 43 months, *see United States v. Ventura*, 650 F.3d 746, 747–49 (D.C. Cir. 2011), and a downward variance of 10 months, *see United States v. Gardellini*, 545 F.3d 1089, 1090 (D.C. Cir. 2008).

161. These data reflect the five most common types of cases to involve appeals of the reasonableness of the sentence imposed. For a variety of reasons, including selection, this may not correlate with the most common types of cases charged and/or types of convictions appealed within the circuit generally. These data do not include sentences imposed for violations of probation or supervised release because there was often no way to separate the crime type of the underlying case from the nature of the supervision violation (e.g., failing a drug test or committing a new-law violation) for the purpose of analyzing case type.

162. If a case involved multiple types of charges or a single charge that could fall into multiple categories (e.g., use of a firearm in conjunction with drug trafficking or money laundering drug proceeds), the author made a subjective judgment about the primary type of crime based on the facts of the case.

163. *See supra* note 119 and accompanying text.

3. *Crime Type Matters*

The five most common types of sentences reviewed in the Third Circuit were in cases involving drugs, guns, child pornography, white-collar, and bank robbery. The type of case appeared to make a significant difference in the likelihood that the Third Circuit would affirm or reverse a variance. The Third Circuit reviewed 10 variances in cases involving drug trafficking, 2 upward variances and 8 downward variances, and the court affirmed all but 1 of them, a downward variance,¹⁶⁴ for a relative reversal rate of approximately 13% of the downward drug variances. By contrast, the Third Circuit reviewed 7 variances in white-collar cases, 6 downward and 1 upward, and the court reversed 4 of the 6 downward variances, for a relative reversal rate of approximately 67% of the downward white-collar variances.

There were only three types of cases in the Seventh Circuit data set for which more than one sentence was reviewed for substantive reasonableness during the study period: drugs, guns, and white-collar crime, but the type of case may make a difference in the results of review between those three categories. In drug and gun cases, the Seventh Circuit affirmed all (100%) of the downward variances, but it reversed 3 of the 6 upward variances, for a relative reversal rate of 50%. By contrast, in white-collar cases, the data set was so small and one-directional (2 downward variances) that the comparative results are not statistically significant. The court reversed 1 of the 2 downward variances that it reviewed, for a relative reversal rate of 50%.

The five most common types of cases reviewed by the Eighth Circuit for substantive reasonableness were: drugs, guns, child pornography, white-collar, and violent crimes. The Eighth Circuit's relative deference to in-guideline sentences made it difficult to detect a case-type difference, although it did show more of a willingness to reverse variance sentences in child pornography cases, up or down, than its overall reversal rate for variances, with a 50% reversal rate for variances in child pornography cases compared with 19% of variances overall.

There were only four types of cases in the Ninth Circuit data set for which more than one sentence was reviewed for substantive reasonableness during the study period: white-collar, drugs, immigration, and violent crimes.¹⁶⁵ While the overall number of reviewed sentences

164. See *United States v. Wells*, No. 06-1456, 2008 WL 2123337, at *1 (3d Cir. May 21, 2008) (reversing Wells's 174-month sentence when his corresponding guideline range was 210–62 months imprisonment).

165. The violent-crime category included bank robbery, terrorism, and arson cases, none of which individually was numerous enough to warrant individual categorization.

from which to draw a case-type conclusion was fairly small, the Ninth Circuit appeared to be significantly more likely to reverse a sentence as unreasonable in cases involving violent crimes. The Ninth Circuit affirmed all (100%) of the variances that it reviewed in white-collar, drug, and immigration cases, but reversed 67% of the variances that it reviewed in cases involving violent crimes, without apparent regard to the size or direction of the variance.

The Tenth Circuit showed a similar split between violent and nonviolent crimes. The five most common types of cases reviewed by the Tenth Circuit for substantive reasonableness were: child pornography, immigration, murder, drugs, and white-collar crimes. Like the Ninth Circuit, the Tenth Circuit affirmed all (100%) of the variances that it reviewed, without regard to size or direction, except in murder cases, in which it reversed 67% of them.

4. *Inconclusive*

The five most common types of sentences reviewed in the Second Circuit were in cases involving drugs, guns, child pornography, immigration, and white-collar crimes. No significant impact by case type was seen in the Second Circuit, possibly because the data set was too small for any meaningful pattern to emerge.¹⁶⁶

The five most common types of sentences reviewed by the Fourth Circuit, by case type, involved drugs, guns, immigration, white-collar crimes, and bank robbery. The Fourth Circuit's strong preference for upward variances (i.e., the tendency to affirm upward variances and reverse downward ones)¹⁶⁷ may have obscured any impact by case type. In the cases involving these five most common types of crimes, the Fourth Circuit affirmed all upward variances (100%) and reversed all downward variances (100%), except in drug cases. The court reviewed 6 cases involving variances of sentences for drug trafficking crimes and affirmed all of them (100%), 4 of which were upward variances and 2 of which were downward variances, but the data set of drug cases (6 total) is too small from which to draw a definitive conclusion about the significance of case type in determining the court's likelihood of reversal.

166. For example, the most common type of case for which a variance sentence was appealed for substantive reasonableness in the Second Circuit was drug cases. The Second Circuit reviewed 10 variances in drug cases, 1 involving an upward variance and 9 involving downward variances. The Second Circuit affirmed the upward variance, affirmed 6 of the downward variances, and reversed 3 of the downward variances. This is consistent with the Second Circuit's overall pattern of relatively high likelihood of reversing downward variances, across case types (27% of the time). See *supra* Part III.C.

167. See *supra* Part III.C.

The five most common types of sentences reviewed by the Sixth Circuit, by case type, involved drugs, guns, immigration, white-collar crimes, and child pornography. The Sixth Circuit's relative deference to district court sentences created a data set of reversals that was too small from which to draw any conclusions about case-type preferences.¹⁶⁸

At first glance, the Eleventh Circuit appeared to have a case-type correlation with the likelihood of reversal, but, when the direction of the variance was also taken into consideration, the results instead seem to be explained by the Eleventh Circuit's preference for upward variances. The five most common case types reviewed for reasonableness by the Eleventh Circuit during the study period were: white collar, guns, immigration, drugs, and child pornography. The Eleventh Circuit affirmed of the variant sentences that it reviewed in gun and immigration cases, while reversing several variances in white-collar, gun, and child pornography cases. This case-type correlation becomes insignificant when the direction of variances imposed by the district courts in the data set is considered. In the cases under review, the district courts, collectively, departed upward in approximately 86% of the gun cases and approximately 81% of the immigration cases that the Eleventh Circuit reviewed for substantive reasonableness. By comparison, the district courts, collectively, departed downward in a majority of the other reviewed cases: in approximately 64% of the reviewed drug cases, approximately 63% of the reviewed white-collar cases, and 50% of the reviewed child pornography cases. The case-type divergence in the direction of variances at the district court (or notice of appeal) level in conjunction with the Eleventh Circuit's preference for upward variances, rather than the circuit court's case-type preference, therefore, seems to account for this variation by case type on review.

168. *See supra* Part III.B. The Sixth Circuit does not show extraordinary deference to the sentencing court, but, of the courts that are identified as retaining appellate discretion to review sentences, the Sixth Circuit did so at the smallest rate (approximately 10% of the time across the board). Once that reversal rate was spread over the 5 most common case types reviewed for sentencing reasonableness in the Sixth Circuit, the resulting individual case-type data sets were too small from which to draw further conclusions. For example, the 2 most common types of sentences that the Sixth Circuit reviewed were those involving white-collar crime and child pornography. The court reviewed 10 variances in white-collar cases, 6 upward variances and 4 downward variances, reversing 2 of the upward variances and affirming the rest. The court reviewed 11 variances in cases involving child pornography, 6 upward variances and 5 downward variances, and reversed 1 of the downward variances, affirming the rest. On the surface, the difference between a 20% and 9% relative reversal rate between the 2 case types may seem meaningful, but it is the difference of only 1 reversal and therefore too insignificant a difference from which to extrapolate any greater correlation.

The data set in the D.C. Circuit was too small from which to draw any conclusions about the effect of case type on the court's review of variances.

H. Narrative Analysis: The Geography of Sentencing

When these various data points are put together, a portrait emerges. Justice Scalia was right. Post-*Booker*, the individual circuits have developed sentencing "personalities." The mismatch of advisory guidelines and unguided reasonableness review has reanimated the lack of uniformity that existed before the adoption of the guidelines. A defendant's treatment, and therefore ultimate sentence, depends on where the crime occurred. What follows is a narrative interpretation of the analyses.

1. Circuits in Which the District Courts Rule

In the First, Fifth, and Tenth Circuits, defendants should really care about their individual sentencing judge at the district court level because all three circuits show extraordinary deference to district court sentences, including variant ones. Sentencing deference in these circuits is exercised at the time of sentencing and not on substantive reasonableness review. This is particularly true in the Fifth Circuit, which has formally adopted a presumption of reasonableness for guideline sentences and has affirmed all of the variances that it reviewed during the course of the study. It is slightly less true in the Tenth Circuit, which, despite having a high rate of affirmance overall, demonstrated discomfort with variances that were particularly large. Together, these circuits appear to engage in exactly the approval of "virtually any sentence within the statutory range that the sentencing court imposes, so long as the district judge goes through the appropriate formalities," exactly that which Justice Scalia predicted in *Booker*.¹⁶⁹

2. Appellate Discretion

At the other end of the spectrum, sentences in the Sixth Circuit are effectively determined in Cincinnati, so defendants should really care about the composition of the three-judge panel that they draw to review their sentences for reasonableness on appeal. The court shows significantly less deference to district court sentences on appeal, and it demonstrates more willingness to reverse sentences that it finds to be substantively unreasonable, without significant consideration of the size or direction of the variance.

169. See *United States v. Booker*, 543 U.S. 220, 312 (2004) (Scalia, J., dissenting in part).

3. *Time Warp: De Facto Mandatory Guidelines*

In the Second, Third, and Eighth Circuits, defendants' sentences are still decided by the United States Sentencing Commission in Washington, D.C., so their biggest concern should be the Commission's promulgation of the guidelines. The Second and Third Circuits, in theory, espouse a willingness to engage in meaningful substantive reasonableness review, but, in practice, they fail to do so. Both courts rejected *Rita's* invitation to presume that in-guideline sentences are reasonable¹⁷⁰ and have even held that sentences within the child pornography guidelines are presumptively unreasonable.¹⁷¹ Nonetheless, when the pattern of individual cases is analyzed, both courts show significantly more deference to in-guideline sentences than to variances on review and appear to be particularly uncomfortable affirming large downward variances.

The Eighth Circuit also remains highly guideline bound, although, in this circuit, that is a conscious doctrinal choice. The court has formally adopted a presumption of reasonableness when reviewing in-guideline sentences,¹⁷² and, consistent with that presumption, shows significantly more deference to in-guideline sentences than to variances on appeal. In keeping with its deference to the guidelines, the size of variances plays a significant role in the court's willingness to affirm them. Not only are in-guideline sentences "safer" on appeal in the Eighth Circuit, but the farther away from the applicable guideline range that the district court strays, the more likely the appellate court is to find the sentence substantively unreasonable.

4. *Severe Sentences*

The Fourth and Eleventh Circuits are just bad places to be sentenced as a defendant. The Fourth Circuit's theory and practice align in its adherence to the guidelines. The court has adopted a presumption of reasonableness for in-guideline sentences. Consistent with that presumption the court shows significantly more deference to in-guideline sentences than to variances. The only time that the court shows a willingness to engage in meaningful sentencing review is when it affirms upward variances, without regard to the size of the variance.

The Eleventh Circuit shows much more overall deference to district court sentences than the Fourth Circuit does, overall, but the Eleventh Circuit is more likely to affirm an upward variance than a downward

170. See *supra* note 29 and accompanying text.

171. See *supra* note 93 and accompanying text.

172. See *supra* note 25 and accompanying text.

one. At the same time, it is particularly likely to reverse large downward variances.

5. *Sentencing Relief*

In contrast, *Booker* appears to have made the Seventh Circuit at least a slightly better place in which to be sentenced as a defendant. The court has also embraced its post-*Booker* discretion, although it is less comfortable affirming larger variances and upward variances of any size.

6. *Unused Discretion*

The Ninth Circuit should fall into the “Appellate Discretion” category, except that in the Ninth Circuit defendants do not seem to be appealing their sentences very often. The court shows significantly less deference to district court sentences on appeal and more willingness to reverse sentences that it finds to be substantively unreasonable, treating in-guideline sentences and variances essentially the same on appeal; however, the circuit has relatively few appeals of guideline sentences (by defendants) for substantive reasonableness.¹⁷³

I. *Issues for Future Study*

The data give rise to several questions outside of the scope of this study and the appeals-level data collected, thus warranting future study: the extent of district-by-district sentencing variations, in particular within the “District Court Deference” circuits identified in this study;¹⁷⁴ the relationship between a particular circuit’s rate of affirming variances, reversing in-guideline sentences, and the number of variances imposed at the district court level (a causation problem);¹⁷⁵ the relationship between a particular circuit’s rate of affirming variances, reversing in-guideline sentences, and the percentage of sentences (in-guideline and variances) that are appealed (a selection problem) (e.g., few defendants appeal their in-guideline sentences as unreasonable in the Ninth Circuit even though the court has shown a real willingness to reverse them); and the relationship between the number/rate of *Anders* briefs within a circuit (particularly in the

173. The data set of cases in the D.C. Circuit is too small to draw any conclusions about sentencing “personality.”

174. See, e.g., Gelacak et al., *supra* note 105, at 360 (finding “large variations in departure practice among district courts *within* circuits” under the pre-*Booker* mandatory guideline scheme).

175. See *id.* at 358 (finding a link, pre-*Booker*, “between a circuit court’s appellate jurisprudence and the actual departure behavior of district judges”).

Fourth Circuit, where that number/rate seems disproportionately high) and that circuit's rate of reversing sentences on appeal. The study also did not consider or address caseload, prosecutorial decisions, the role of plea or cooperation agreements, or disparities at the district court level (with the possible exception of those created or driven by the appellate-review practice of the circuit in which a particular district is located), all of which may provide alternate explanations of some of the initial narrative impressions of the circuits.

V. PRELIMINARY IMPLICATIONS FROM THE RESULTS

A. *Conclusions*

Tens of thousands of people are convicted of federal crimes each year.¹⁷⁶ Complete information about the operation of the federal courts post-*Booker* is essential to improving the fairness and effectiveness of their functioning. These data have a great deal of political and practical importance.

Other studies have documented that, since *Booker*, the average length of sentences for most offenses has not changed significantly and most judges have continued to impose in-guideline sentences.¹⁷⁷ This study demonstrates that the result of the *Booker/Fanfan, Gall/Kimbrough*, and *Rita* line of cases is that federal sentences are determined more by individual district and judge than by an individualized evaluation of each sentence for reasonableness at the appellate level, resulting in disparities between both similarly situated defendants who receive radically different sentences and differently situation defendants who receive similar ones because they are sentenced in different districts/circuits. In other words, two defendants committing the same crime under the same circumstances may receive very different sentences depending on the sentencing idiosyncrasies of each defendant's particular judge. Meaningful appellate review of sentences is critical to assure that there is a proper balance between avoidance of

176. See generally U.S. SENTENCING COMM'N, OVERVIEW OF FEDERAL CRIMINAL CASES: FISCAL YEAR 2015 (2016), http://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2016/FY15_Overview_Federal_Criminal_Cases.pdf.

177. See U.S. Sentencing Comm'n, 2009 ANNUAL REPORT tbl. N (2009), <http://www.ussc.gov/about/2015-annual-report/archive/annual-report-2009> (noting that in Fiscal Year 2009, 56.8% of federal sentences were imposed within the applicable guideline range, 25.3% were the result of a government-sponsored downward variance, and only 17.9% were imposed outside of the applicable guidelines range without the government's motion). To compare, prior to *Booker*, defendants received in-guideline sentences in approximately two-thirds to three-fourths of cases. See U.S. Sentencing Comm'n, SOURCEBOOK OF FEDERAL SENTENCING STATISTICS, ANNUAL REPORT 2004 49 fig. G (2004) (noting percentages of in-guideline sentences from 1992–2004 between 64–78%).

unjustified disparities and recognition of the importance of individualizing sentences to reflect the totality of the circumstances regarding an offender and their offense.

The Supreme Court's remedial opinion in *Booker* and its progeny have left the federal circuit courts of appeal with two separate but interrelated options: (1) whether to engage in a formal presumption of the substantive reasonableness of in-guideline sentences under *Rita*;¹⁷⁸ and (2) whether to engage in meaningful or purely formal ("rubber stamp") review of the substantive reasonableness of a sentence in individual cases regardless of whether such sentences are within or outside the applicable guideline range.¹⁷⁹ The result is a tripartite division among the circuits.

The first category of circuits comprises circuits that employ a presumption of substantive reasonableness, resulting in a rubber stamp of in-guideline sentences and a near-universal rejection of sentencing variances.¹⁸⁰ The second category of circuits comprises those that do not employ a presumption of substantive reasonableness for in-guideline sentences but rather rubber stamp all sentences, so long as the district court indicated that it understood the discretion that it was exercising, which has increasingly become conflated with a review of procedural reasonableness. The third category of circuits neither employs a presumption of substantive reasonableness for in-guideline sentences nor rubber stamps any sentences.¹⁸¹

Only sentences within the third category of circuits comply with the Sixth Amendment right to trial by jury and the Fifth Amendment rights to proof beyond a reasonable doubt and a grand jury indictment. Only these sentences embody the advisory-guideline system envisioned by the Court in *Booker* as a remedy for its constitutional holding.

Circuits that employ a *de jure* presumption of reasonableness for in-guideline sentences and a *de facto* presumption of unreasonableness for variant sentences have created a circuit-wide *de facto* mandatory

178. *United States v. Booker*, 543 U.S. 220, 262–65 (2005).

179. *Id.*

180. *See, e.g.*, Carissa Byrne Hessick & F. Andrew Hessick, *Rita*, *Claiborne*, and *the Courts of Appeals' Attachment to the Sentencing Guidelines*, 19 F. SENT'G REP. 171, 174 (2007) (noting "a larger pattern of resistance by the courts of appeals to the Supreme Court's efforts to reconstruct the law of sentencing" in *Booker*).

181. While it is theoretically possible to have a fourth category of circuit courts—ones that employ a presumption of reasonableness for in-guideline sentences while rubber stamping sentences below or above the guideline range—given the philosophical inconsistency between rubber stamping and employing a strong presumption of guideline sentences, there are presently no circuits that this Article would categorize in this way.

guideline regime, in violation of the Fifth and Sixth Amendments and *Booker* and *Fanfan*.

Circuits that employ no presumption of reasonableness, but simply rubber stamp all district court sentences, violate the Fifth and Sixth Amendments in a more subtle way. By affirming essentially all district court sentences so long as they are procedurally reasonable, these circuits have created a *de facto* mandatory guideline system within some of their districts or even individual courtrooms—namely, those districts and courtrooms in which judges, while articulating the knowledge that they are permitted to impose a below-guideline sentence never actually do so, but nonetheless have their guideline sentences routinely affirmed at the circuit level. The “excessive sentencing disparities”¹⁸² that result from the luck of the draw of individual sentencing judge (judge that will vary from the guideline when appropriate v. judge who always imposes guideline sentences) in these circuits is similarly unconstitutional.

B. Limitations

This study suffers from two distinct limitations. First, it is based entirely on data from reported appellate-court opinions (i.e., cases in which a district court’s sentencing decision has been challenged and the appellate court has published its decision reviewing the sentence). In addition to excluding contested sentences that are not appealed (and perhaps some contested sentences that are appealed but whose appeals are not reported), it also excluded most sentences that are imposed after a negotiated guilty plea.¹⁸³ This latter exclusion is particularly significant in the federal system, because the federal rules of appellate procedure generally prohibit judges from agreeing in advance to impose negotiated sentences.¹⁸⁴ Second, it involves the collection of data for only a short period of time (approximately four years) immediately following the Court’s decisions in *Gall* and *Kimbrough*.

C. Additional Validation

This study has provided a number of avenues for future research regarding the sentencing consequences of *Booker*. One such avenue would be to repeat the study over a longer period of time, further

182. *Booker*, 543 U.S. at 312 (Scalia, J., dissenting in part).

183. Plea agreements generally require a waiver of the defendant’s right to appeal the sentence imposed.

184. See FED. R. CRIM. P. 11(c) (prohibiting the sentencing judge from participating in plea negotiations).

removed from the Court's decisions in *Gall* and *Kimbrough*, to see if the correlations that this study found continue to hold true.

A second avenue of future research would be qualitative research involving interviews with individual judges about their decision-making processes under the new advisory-guidelines system. Such qualitative results could support the quantitative finding of this study that, in some districts and for some judges, the guidelines are still de facto mandatory.

A third avenue for future research would be to replicate this study at the district court level (i.e., using sentencing data from sentences that do not get appealed and/or publicly reported). The findings of this study may be skewed by the selection of sentencing decisions for appeal.

D. Final Thoughts

Disparities in sentencing may never be wholly eliminated, but one of their major causes, the tension between sentencing discretion at the district court level and substantive-reasonableness-review discretion at the circuit court level, can and should be eliminated. Of course, the results of this study will only have meaning if they are actually used by Congress and the federal judiciary to make more uniform and fair sentencing decisions.

No guideline system can fully account for the circumstances that will produce sentencing disparities, and by removing or severely limiting judicial discretion in sentencing, judges cannot adjust and correct accordingly. A rigid system intended to reduce overall disparity actually creates disparities different than, but equally unwarranted as, those documented in this Article. Mandatory guidelines are overly blunt instruments, bringing undue focus upon some factors (like criminal histories and drug quantities) to the exclusion of other important considerations, including the defendant's role in the offense, the presence or absence of violence, the risk of recidivism, and the personal characteristics of the defendant. The goal of this study, therefore, is to provide both the courts and the public with accurate information so that they can examine whether justice is being achieved.¹⁸⁵

185. The alternative to rendering the USSG advisory, which was preferred by the four justices dissenting from the Court's remedial opinion in *Booker* (also constituting four-fifths of the merits majority), would have been to require the prosecution to plead and prove aggravating sentencing factors to the jury as elements of the offense under *Apprendi*.

